Analysis of Sox10 Target Genes in Zebrafish Early Development

Chipperfield, Tom

Award date:
2009

Awarding institution:
University of Bath

Analysis of *Sox10* target genes in zebrafish early development.

By Thomas Richard Chipperfield

A Thesis Submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

August 2009

Author Signature . . . . . .

Abstract

The neural crest is a transient population of cells that forms a diverse range of derivatives in vertebrate embryos. Neural crest cells also migrate extensively throughout the embryo. The specification of a number of neural crest derivatives, including pigment cells and neurons and glia of the peripheral nervous system, is dependent on the transcription factor Sox10. In *sox10* mutant zebrafish embryos, these neural crest derivatives fail to specify and subsequently the cell differentiation and migration fails leading to apoptosis. *Sox10* mutant embryos also display an ear defect although the precise role of Sox10 in the ear is less well defined. Thus Sox10 controls an extensive gene regulatory network that drives the development of an important subset of neural crest derivatives and also functions during ear development. This gene regulatory network is currently poorly defined.

The aim of this project was to identify genes that are both direct and indirect targets of Sox10 to further elucidate this gene regulatory network. To achieve this, a microarray approach was adopted. Initially, fluorescence activated cell sorting was employed to enrich for *sox10* expressing cells from 24 hours post fertilization *sox10*:GFP transgenic embryos. The transcriptomes of WT and *sox10* mutant cells were compared by microarray analysis to identify differentially regulated genes. A large number of target genes were identified by this method and by an unbiased *in situ* hybridization screen, 28 genes were validated. Of these, 23 genes were expressed in cells of the neural crest and down-regulated in *sox10* mutant embryos. The majority of these genes were expressed in cells of the melanocyte and xanthophore lineages. 5 genes were expressed in the ear (otic vesicle) of which three otic vesicle genes were down-regulated while two otic vesicle genes were up-regulated in *sox10* mutant embryos. Unfortunately due to time constraints, a study into the function of one of these target genes could not be completed.

The series of validated genes identified during this project has opened new opportunities for research and has identified a number of highly expressed marker genes that will be useful in future studies. In addition, the microarray data presented will be a useful resource to aid the identification of further targets of Sox10.

Table of Contents

Table of Figures

List of tables

List of Abbreviations

7 –AAD	7-Amino-actinomycin
µg	microgram
µl	microlitre
µM	micromolar
bp	base pair
ANOVA	Analysis of Variance
ATP	Adenosine Triphosphate
BCIP	5-Bromo-4-chloro-3-indolyl phosphate, toluidine salt
BMP	Bone Morphogenic Protein
BSA	Bovine Serum Albumin
CARB	Carbenicillin
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
CMPD	Capomelic Dysplasia
CNS	Central Nervous System
cRNA	complementary RNA
DEPC	Diethyl Pyrocarbonate
DIG	digoxigenin
DNA	Deoxyribonucleic Acid
DRG	Dorsal root ganglia
E. coli	*Escherichia coli*
ECM	Extracellular matrix
EDTA	ethylenediaminetetraacetic acid
EGF	Epidermal Growth Factor
EMT	Epithelial to mesenchymal transition
ENS	enteric nervous system
ENU	*N*-ethyl *N*-nitrosourea
EPI-NCSC	Epidermal Neural Crest Stem Cell
EtBR	Ethidium Bromide
EtOH	Ethanol
FACS	Fluorescence Activated Cell Sorting
FBS	Fetal Bovine Serum
FDR	False Discovery Rate
FGF	Fibroblast Growth Factor
FWER	Family Wise Error Rate
GDNF	glial cell line derived neurotrophic factor
GEPAS	Gene Expression Pattern Analysis Suite
GFP	Green Fluorescent Protein
GGF	Glial growth factor
GMP	Guanine monophosphate
GO	Gene Ontology

GTP	Guanosine Triphosphate
HM	Hybridization mix
HMG	High mobility group
hpf	hours post-fertilization
IPTG	Isopropyl-beta-D-thiogalactopyranoside
ISH	*in situ* hybridization
Kb	1,000 base pairs
LB	Luria Broth
M	molar
MeOH	Methanol
Mg	milligram
MHB	mid-brain/hind-brain boundary
Ml	millilitre
Mm	millimetre
mM	millimolar
MM	Mis-match probe (see probe pair)
mRNA	messenger Ribonucleic Acid
MSC	melanocyte stem cell
NBT	Nitro blue tetrazolium chloride
NC	neural crest
NCC	neural crest cell
NCSC	neural crest stem cell
ng	nanogram
nm	nanometre
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PCWH	peripheral demyelinating neuropathy, central dysmyelinating leukodystrophy, Waardenburg syndrome and Hirschsprung disease
PFA	Paraformaldehyde
pg	picogram
pH	Measure of acidity or basicity of a solution
PLL	posterior lateral line primordium
PM	perfect match probe (see probe pair)
PNS	Peripheral Nervous System
PTU	1-phenyl-2-tiourea
QRT-PCR	Quantitative RT-PCR
RMA	Robust multi-array analysis
RNA	Ribonucleic Acid
RNase	Ribonuclease
RPE	retinal pigment epithelium
rpm	Revolutions per minute
RT-PCR	Reverse Transcription PCR
SAGE	Serial Analysis of Gene Expression
siRNA	small interfering RNA
SKP	Skin derived precursor
TF	Transcription factor
Tg	Transgenic
TGF	Transforming growth factor
T_m	Melting Temperature
UV	ultraviolet
V	volt
V-ATPase	vacuolar-H^+ ATPase
WS	Waardenburg syndrome
WT	Wild Type
w/v	weight/volume
XMP	xanthosine monophosphate
ZFIN	zebrafish information network (www.zfin.org)

Acknowledgements

Firstly, thank you to my supervisor Dr. Robert Kelsh, without his continuous enthusiasm and support this work would not have been possible. I owe him a great debt of gratitude for the opportunity to work in his lab. I would also like the thank all the members of the Kelsh and Ward “zebramouse” labs, past and present, for always helping to share the problems. Thanks also to our fish technician Marc Sheddon for his expert fish care. I would also like to thank all my friends and my family for their continuous support in so many ways.

A number of people contributed directly to this project. Amy Boyd initiated the project and thanks to her for teaching me the workings of the FACS machine and many other techniques in the laboratory. Thanks also to all the students who contributed to this project during their time in our lab although particular thanks must go to Caroline Nelson for her hard work. A huge thank you to all the staff at the UCI DNA and Protein Microarray facility, especially Cherryl Nugas Selby, Seung-Ah Chung and Denis Heck, for not only processing the RNA and performing the microarray hybridization experiments but also for their kind help and advice that enabled us to gain a much stronger understanding of microarray work.

Chapter 1: Introduction

1.1 Introduction to the Neural Crest

1.1.1 Overview

The neural crest (NC) is a critical vertebrate embryonic cell population sometimes referred to as the fourth germ layer. Neural crest cells (NCCs) arise at the border between neural and non-neural ectoderm and after neurulation lie at the dorsal aspect of the neural tube. Subsequently, NCCs undergo an epithelial to mesenchymal transition (EMT) and delaminate from the neural tube before migrating to numerous locations within the developing embryo. The initially multipotent population of NCCs differentiates into a diverse variety of cell types including pigment cells, neurons and glia of the peripheral nervous system and craniofacial cartilage. NCCs have thus provided an excellent model for the study of cell specification, differentiation and migration. In addition, the properties of NCCs have strong parallels with both stem cells and invasive cancerous cells. A number of human diseases are also attributed to errors in NC development, such diseases are termed neurocristopathies. Therefore study of the NC has been driven by both their key developmental importance and their relevance to human disease aetiology.

1.1.2 Formation and Induction of the Neural Crest

The neural crest forms at the lateral edges of the neural plate at the boundary with the prospective epidermis. As neurulation progresses, NCCs are positioned at the tip of the neural folds and are found in the dorsal NT after NT closure. In zebrafish there is no folding of the neural plate but instead the neural keel is formed by a thickening of the neural plate. The neural keel then separates from the prospective ectoderm and subsequently NC mesenchymal cells can be distinguished in the dorsal portion of the neural keel (Eisen and Weston, 1993). As with other vertebrate species, the development of NCCs occurs in a rostrocaudal progression (Raible et al., 1992).

Several major signalling pathways converge to induce formation of the NC. In accordance with the classical gradient model, the NC is induced from ectoderm at intermediate levels of Bone Morphogenic Protein (BMP) signalling, with neural tissue induced at high levels of BMP signalling (Marchant et al., 1998). Analysis of zebrafish mutants has supported this model; the *swirl*/*bmp2* zebrafish mutant abolishes BMP signalling and displays no induced NC while a mutant that has reduced BMP signalling, for example *somitabun*/*smad5*, displays an expanded zone of induced NC (Nguyen et al., 1998). However, other studies have identified that an intermediate level of BMP signalling is necessary but not sufficient to induce NC. Induction of NCCs from *chordin* treated *Xenopus* neural plate explants required additional signals, in particular Wnt or Fibroblast

Growth Factor (FGF) signalling (LaBonne and Bronner-Fraser, 1998). Such findings have led to the development of the double gradient model which stipulates that intermediate BMP signalling in conjunction with other signalling molecules including Wnts and FGFs from surrounding tissues are required to induce the NC (Aybar and Mayor, 2002). Loss of function experiments in zebrafish has also identified a key role for Wnts during NC induction. Using a heat inducible transgenic zebrafish embryo, in which canonical Wnt signalling could be blocked, the authors identified a key period when inhibition of Wnt signalling prevented the induction of NCCs (Lewis et al., 2004). In response to the correct combination of signals, a set of transcription factors termed neural plate border specifiers (including *zic1, msx* genes, *dlx* genes and *pax* genes) are up-regulated and define the embryonic region capable of forming the NC (Sauka-Spengler and Bronner-Fraser, 2008). Neural plate border specifier genes include *msx1* which is a known target of BMP signalling (Tribulo et al., 2003). *Msx1* is capable of inducing expression of the NC specifier genes *slug*, *snail* and *foxd3* in *Xenopus* embryos (Tribulo et al., 2003). MSX1 also up-regulates and cooperates with PAX3 to induce NC specifier gene expression by mediating WNT and FGF signals (Monsoro-Burq et al., 2005). Thus in response to the neural plate border specifier genes, early NC genes known as NC specifier genes are up-regulated (Sauka-Spengler and Bronner-Fraser, 2008).

1.1.3 Neural Crest Specification and Differentiation

The NC specifier genes are a group of transcription factors and are used as early NC markers. This set of genes includes *snail* genes, *soxE* genes, *foxd3, tfap2, twist, c-Myc* and *Id* genes (Sauka-Spengler and Bronner-Fraser, 2008). The combination of these genes distinguishes NCCs from other neuroepithelial cells and initiates further steps in NC development such as epithelial to mesenchymal transition (EMT), delamination, specification of NC derivatives, cell migration and cell survival. Evidence so far shows that the NC specifier genes display extensive co-ordination and cross-regulation in a gene regulatory network. For example, the depletion of *Sox10* by morpholino knockdown in *Xenopus* embryos inhibited the expression of *snai2* and *foxd3* (Honore et al., 2003). Although, the expression levels of *snai2* and *foxd3* are not affected in the early NCCs of zebrafish *sox10* mutants (Lopes et al., 2008). Over expression of *snail, slug* and *sox9* all increased the expression of *sox10* (Aoki et al., 2003, Honore et al., 2003). The NC specifier gene network ensures that the NCC state is maintained and downstream NC effector genes are activated (Meulemans and Bronner-Fraser, 2004).

The potential of a cell is defined as the fate or fates that can be adopted by that cell or its progeny given appropriate environmental conditions. A cell is therefore multipotent if it has the potential to generate more than one distinct cell type (Kelsh,

2006). It should be noted that the fate a cell actually adopts is not necessarily the same as what it could do given appropriate signals. Single cell labelling and subsequent cell lineage analysis has identified, in a variety of model organisms, that NCCs are initially multipotent (Bronner-Fraser and Fraser, 1989, Collazo et al., 1993, Serbedzija et al., 1994, Raible and Eisen, 1994). Some NCCs were capable of producing more than one type of derivative while some NCCs only produced one derivative. As such, NCCs are not only multipotent but the NCC population is heterogeneous. How a wide range of NC derivatives are specified from a multipotent precursor population is thus a key question. Two options exist for NCCs to achieve this; cells could be directly specified from the multipotent precursor population, or cells could undergo progressive fate restriction. Currently evidence supports cells progressively losing potential, passing through a series of partially restricted precursors before finally specifying and subsequently committing to a single fate (Kelsh, 2006). For example, *Sox10* mutant embryos only display defects in a subset of NC derivatives, the non-skeletogenic (non-ectomesenchymal) derivatives which include pigment cells and the neurons and glia of the peripheral nervous system (PNS). The skeletogenic derivatives, which include cartilage and fin mesenchyme, are unaffected (Southard-Smith et al., 1998, Dutton et al., 2001a). This indicates that the multipotent NC has been restricted into skeletogenic and non-skeletogenic precursor populations. A restriction between neuronal and pigment cell precursors has also been identified (Henion and Weston, 1997). It should also be noted that NC specification and fate choice occurs with temporal and spatial differences. Zebrafish trunk NCCs that are specified to a neuronal fate appear first while pigment cells appear later (Raible and Eisen, 1994). NCCs at different axial levels produce different cell types, for example, only vagal NCCs produce smooth muscle that contributes to the heart (Li et al., 2003) while NCCs at all axial levels produce neurons, glia and pigment cells.

A cell becomes specified to a particular fate when it starts to express characteristics of a particular cell type or fate, but it should be noted that a specified cell is not necessarily committed to a particular fate (Kelsh, 2006). Thus a specified cell will express key molecular markers indicative of the future cell type. The earliest marker of melanocytes is the transcription factor *Mitf* (the zebrafish homolog is *mitfa*). *Mitf* is considered to be the key gene in the switch from a bipotent glial/melanocyte precursor to a melanocyte and is the master regulator of melanogenesis (Opdecamp et al., 1997). *Mitf* expression requires co-regulation by SOX10 and PAX3, thus two NC specifier genes activate this master switch gene (Potterf et al., 2000, Elworthy et al., 2003). In addition, activation of the Wnt signalling pathway by injection of mRNA into single cells promotes the formation of melanocytes in zebrafish embryos (Dorsky et al., 1998). During the same study, injection of dominant negative Wnt1 mRNA to inhibit Wnt signalling promoted neuron formation at the expense of melanocytes. It has been shown that the *mitfa*

promoter contains Wnt signalling responsive binding sites (Dorsky et al., 2000). Thus there is a key requirement for both TF expression and signalling by extrinsic instructive cues for correct NCC specification. Subsequently *Mitf* regulates the expression of a number of melanocyte specific genes (Murisier and Beermann, 2006). Thus the model of NC specification and differentiation presented here involves NC specifier genes activating master regulator transcription factor genes, in concert with environmental signals, to activate genes responsible for NCC differentiation, migration and survival (Dutton et al., 2001a). The NC specifier transcription factors therefore activate molecular cascades necessary to initiate further NCC development. Examples of master regulator transcription factors that are key to the development of a particular NC cell type include, *phox2b* (enteric nervous system) and *ngn1* (sensory neurons) in mouse (Pattyn et al., 1999, Perez et al., 1999) and zebrafish (Elworthy et al, 2005, Carney et al, 2006).

1.1.4 Epithelial to Mesenchymal transition and Crest cell migration

NC specifiers regulate some of the characteristic features of NCCs, namely the epithelial to mesenchymal transition (EMT), delamination and migration of NCCs. After specification, NCCs are located within the dorsal neural tube and subsequently transform from epithelial cells into mesenchymal cells, which have migratory properties. NCCs then delaminate from the neural tube. During these events NCCs undergo a series of changes including a loss of apico-basal polarity, dissolution of tight junctions, a change in cell adhesion and they acquire invasive properties which confers the cells with the ability to delaminate and migrate (Sauka-Spengler and Bronner-Fraser, 2008). It should be noted that cancer cells also undergo EMT during tumour progression, thus NC biology relates to cancer biology in this regard (Thiery and Sleeman, 2006). NCCs then undertake extensive migration on stereotypical pathways to various locations within the developing embryo. In mammals, melanocytes migrate on the dorsolateral pathway between the epidermis and the somites while cells of a neural fate migrate on the ventral pathway through the anterior half of the sclerotome (**Figure 1**). The situation is slightly different in zebrafish embryos. Neural fated cells are restricted to the segmentally organized medial (ventral) migration pathway while melanocytes can migrate on both the medial and lateral (dorsolateral) pathways (Raible and Eisen, 1994, Kelsh, 2004). Xanthophores are restricted to migrating on the lateral pathway while iridophores migrate solely on the medial pathway (**Figure 1**) (Kelsh, 2004). Temporal differences in cell migration have also been observed, neural fated cells migrating first on the medial pathway (16 - 18 hpf) and melanocyte fated cells initiating migration slightly later and continuing for a longer period of time (16.5 – 23.5 hpf) (Raible and Eisen, 1994). Migration on the lateral pathway initiates 4-5 hours later than migration on the medial pathway (Raible and Eisen, 1994).

NCCs are initially part of the epithelial NT and have a typical apical-basal polarity. This polarity is lost as cells progress through EMT and gain the ability to migrate. As part of this process tight junctions are lost. This is achieved by the down-regulation of constituent components of tight junctions. For example, *occludin*, a major structural component of tight junctions, is down-regulated in the dorsal neuroepithelium during NT closure (Aaku-Saraste et al., 1996). However, NCCs were not specifically examined during this study. Over expression of the NC specifier *snail* induces EMT in cultured mouse epithelial cells; this process occurred with a concurrent down-regulation in the tight junction proteins *occludin* and *claudin* (Ikenouchi et al., 2003). During NT closure tight junctions are lost as part of the EMT and there is a shift to gap junctions (Shook and Keller, 2003). Gap junctions are formed when two connexons or hemi-channels in neighbouring cell membranes are opposed to allow the passage of small molecules between cells. Connexons are constructed of a hexamer of proteins encoded by the connexin family of genes. Connexons can consist of one type of connexin or a combination of different connexins (Meşe et al., 2007). The function of one connexin, *Cx43*, has been well studied in the NC and is known to modulate the motility and migration rate of cardiac NCCs (Huang et al., 1998, Xu et al., 2001). It appears that the function of *Cx43* is dependent on interactions with the actin cytoskeleton and is independent of gap junction function (Xu et al., 2006). Despite the considered importance of the transition from tight junctions to gap junctions during EMT in both NC and cancer biology, there is little published data regarding the expression of other members of the connexin family during this process and therefore little is known about the function of connexins during EMT and NC migration.

A change in cell to cell adhesion presents another key step that occurs during EMT to infer NCCs with the capacity to migrate. Cadherins are a family of calcium dependent cell adhesion molecules which are split into several groups. The classical cadherins are split into two types, type I includes epithelial (E) and neuronal (N) cadherins while type II covers cadherin-5 and beyond (Halbleib and Nelson, 2006, Taneyhill, 2008). Type I cadherins contain an HAV tri-peptide motif in their extracellular domain that mediates strong cell to cell adhesion and are expressed in stable epithelial cell assemblies, this adhesion motif is lacking in type II cadherins (Halbleib and Nelson, 2006). A switch from the expression of type I cadherins to type II cadherins with the associated change in cell adhesion is correlated with the progression of EMT and acquisition of cell motility. Type I cadherins are down-regulated during EMT by NC specifier genes such as *Foxd3* and *Snail2* (Cheung et al., 2005, Cano et al., 2000). Conversely, type II cadherins are up-regulated in migrating NCCs by NC specifier genes (Cheung et al., 2005). To facilitate migration, a cell not only has to modulate cell to cell adhesion but also cell to extracellular matrix (ECM) adhesion. The integrin family of receptors, that recognise ECM

molecules like fibronectin, represent a major effector of cell to ECM adhesion. Several integrins are expressed by NCCs (Testaz et al., 1999). Loss of function studies has shown that integrins are required for NCC migration both *in vitro* and *in vivo* (Kil et al., 1996). Thus regulation of cell adhesion plays a key role in facilitating the migration of NCCs.

Proteins called matrix metalloproteases (MMPs), which digest the ECM, have also been implicated in both NCC EMT and migration. Interestingly, MMPs are also strongly linked with cancer cell invasion (Kuriyama and Mayor, 2008). The NC specifier gene, *snail*, was shown by microarray analysis to regulate the expression of *MMP-2* in a melanoma cell line. Comparison of cancer cells treated with *snail* antisense cDNA against untreated cells identified a number of important EMT genes including *MMP-2* as down-regulated in treated cells (Kuphal et al., 2005). *MMP-2* is expressed in NCCs during EMT and cell migration but expression is down-regulated in cells that have reached their final destination (Duong and Erickson, 2004). In addition, MMP chemical and MMP-2 morpholino inhibition prevented EMT and NCC dispersion from the NT *in vivo* and *in vitro* (Duong and Erickson, 2004). Members of the metalloprotease/disintegrin ADAM family of genes have also been implicated in NCC migration. *Xenopus* NCC grafts of ADAM-13 protease defective cells did not migrate in WT embryos (Alfandari et al., 2001). Thus inhibition of a range of genes that remodel the ECM results in defective NCC migration.

A NCC must acquire a range of molecular characteristics to confer the cell with the ability to migrate. Subsequently, a number of molecules have been identified that segregate cells onto the NC migration pathways and direct cells to specific final locations, including *ephrins, semaphorins, slit/robo* signalling and non-canonical Wnt signalling (for a review see Kuriyama and Mayor, 2008). Semaphorins are ligands involved in both cranial and trunk NCC migration; neuropilins are the receptors of semaphorins. In mice, NCCs express *Npn2* while migrating through the anterior region of the somite (sclerotome). The repulsive ligand of this receptor, *Sema3f,* is expressed in the posterior region of the somite (Gammill et al., 2006). In *Npn2* mutant mice NCCs were observed to migrate as a single sheet rather than in segmentally arranged streams thus *semaphorin/neuropilin* signalling appears to play a role in excluding migrating NCCs from the posterior area of somites (Gammill et al., 2006). *Ephrins* are also thought play a role in regulating both cranial and trunk NCC migration and in restricting NCCs to the anterior of each somite. The ligand *Ephrin-B1* is expressed in the posterior part of the sclerotome while NCCs migrating through the anterior part of the somite express the receptor *EphB3* (Krull et al., 1997). A number of studies have identified similar complementary *ephrin* ligand and receptor expression patterns (Kuriyama and Mayor, 2008). Addition of soluble Ephrin-B1, which can occupy the receptor without activating it, to chick whole trunk explants, disrupted the segmental migration pattern of NCCs with cells migrating in both anterior and posterior segments of the somite (Krull et al., 1997). *Ephrins* also play a role in

restricting the entry of early NCCs to the medial pathway. NCCs specified as neurons or glial cells are prevented from entering the dorsolateral pathway, when soluble inactive Ephrin-B1 was applied to chick whole trunk explants, again the segmental migration pattern was lost but cells normally restricted to the ventromedial pathway were seen migrating on the dorsolateral pathway (Santiago and Erickson, 2002). It was noted that Ephrin-B receptors were still expressed by melanoblasts during migration on the dorsolateral pathway. It was identified that melanoblast migration was actually stimulated by Eph receptor activation (Santiago and Erickson, 2002). Thus Eph signalling can inhibit the migration of some NCCs while stimulating the migration of other NCCs. Currently no work has been performed to examine if this holds true in zebrafish where melanocytes migrate on both the lateral and medial pathways (Kelsh et al., 2009). The Robo receptor and the Slit secreted ligand represent a third class of molecules involved in NC migration. Early migrating NCCs express Robo receptors and Slits are expressed in the dermomyotome close to the dorsolateral migration pathway. Slit/Robo signalling was observed to repel NCC migration both *in vitro* and *in vivo* to prevent NCCs of a neural fate migrating on the pigment cell pathway (Jia et al., 2005). Slit/Robo signalling has also been implicated in preventing trunk NCCs from migrating into the gut, which is only colonised by vagal NCCs that lack a Robo receptor (De Bellard et al., 2003). Only membrane bound Slit was capable of inhibiting trunk NCC migration, surprisingly soluble Slit, during *in vitro* experiments, increased the motility of trunk NCCs (De Bellard et al., 2003). Thus, Robo/Slit and Ephrin signalling can both have positive and negative effects on NCC migration. Generally though, signals identified with roles in NCC migration thus far have been inhibitory, denying NCCs access to defined areas.

While the majority of NCC migratory signals identified thus far tend to restrict NCCs from a particular area or route, some positive cues that guide cells to a specific location have now been elucidated. The CXCR4-SDF1 signalling system is known to guide neurons and primordial germ cells to their correct positions (Knaut et al., 2003, Miyasaka et al., 2007). In zebrafish, the chemokine *sdf1a* marks the path of the posterior lateral line primordium (PLL). Cells of the PLL also express the *sdf1a* receptor *cxcr4a* (*cxcl12*). Knockdown of expression of either the ligand or the receptor prevents the PLL from migrating (David et al., 2002). The chemokine *sdf1a* has also been shown to be a positive attractant for melanocytes in zebrafish (Svetic et al., 2007). The zebrafish *choker* mutant is a *you*-type muscle mutant; mutant embryos display “u” rather than “v” shaped somites (van Eeden et al., 1996). The *choker* mutant also displays an accumulation of melanocytes in an ectopic collar position and an absence of melanocytes in the lateral stripe. These pigment phenotypes were shown to be a secondary consequence of the muscle phenotype by transplantation experiments (Svetic et al., 2007). Aberrant *sdf1a* expression corresponded to the melanocyte pattern defects; *sdf1a* expression along the

lateral stripe was lost while *sdf1a* was ectopically expressed in the collar region. In addition, human SDF1 soaked beads attached to the epidermis of zebrafish embryos acted as a chemo-attractant to melanocytes (Svetic et al., 2007). Expression of the *sdf1a* receptor, *cxcr4a*, could not be identified in melanocytes, thus the complete molecular mechanism of this chemo-attractant system remains to be fully understood. Glial cell line derived neurotrophic factor (GDNF) may act as a chemo-attractant of vagal NCCs fated to form elements of the enteric nervous system (ENS) (Young et al., 2001, Natarajan et al., 2002). These cells express the GDNF receptor, RET, and GDNF is expressed in a temporal and spatial pattern consistent with a role in guiding ENS cells to appropriate locations. Additionally, *in vitro* experiments demonstrated that ENS neurons grow towards a source of GDNF and loss of RET function resulted in ENS NCCs migrating at a reduced rate (Young et al., 2001, Natarajan et al., 2002). Despite very few positive migration cues having been identified for NCCs thus far, NCCs do migrate in a highly directed fashion to precise locations. It seems likely that further positive migration cues will be discovered including their corresponding receptors.

Migrating cells are highly polarised, extending many projections (lamellipodia and fillopodia) in the direction of travel (Kuriyama and Mayor, 2008). This cell polarisation may play a role in ensuring NCCs migrate in the correct direction. The non-canonical Wnt signalling pathway has been implicated as a key player in regulating NCC polarity during migration. *Wnt11* and *Wnt11r* have been shown to be expressed adjacent to the NC, cells of which express the putative Wnt11 receptor *Frz7* at the time of EMT and initiation of cell migration (De Calisto et al., 2005, Matthews et al., 2008). Loss of function studies of components of the non-canonical Wnt signalling pathway also results in NCCs migration failing (De Calisto et al., 2005, Matthews et al., 2008). NCCs were attracted to grafts of ectoderm over expressing Wnt11 positioned adjacent to NC migration pathways (De Calisto et al., 2005). Analysis of cultured NCCs identified that non-canonical Wnt signalling helped to stabilise the lamellipodia that favour the leading edge of a migrating cell (De Calisto et al., 2005). *Wnt11* and *Wnt11r* both promote NCC migration but are expressed adjacent to but on opposite sides of the NC therefore these molecules are unlikely to function simply as a chemo-attractant (Matthews et al., 2008). Thus although correct non-canonical Wnt signalling by both molecules is essential for migration, full understanding of this process has yet to be deduced. In all of the above studies, non-canonical Wnt signalling was shown to function non-cell autonomously; NCCs did not express *Wnt11* or *Wnt11r*. In contrast, *Wnt11r* was found to function cell autonomously during NCC EMT and was essential for correct formation of the dorsal fin (Garriock and Krieg, 2007). In *Wnt11r* morphant embryos, skeletogenic NCCs and muscle cells that contribute to the dorsal fin failed to migrate to this location. Therefore non-canonical Wnt

signalling and the role that it plays in NCC migration displays an additional level of complication that needs to be examined.

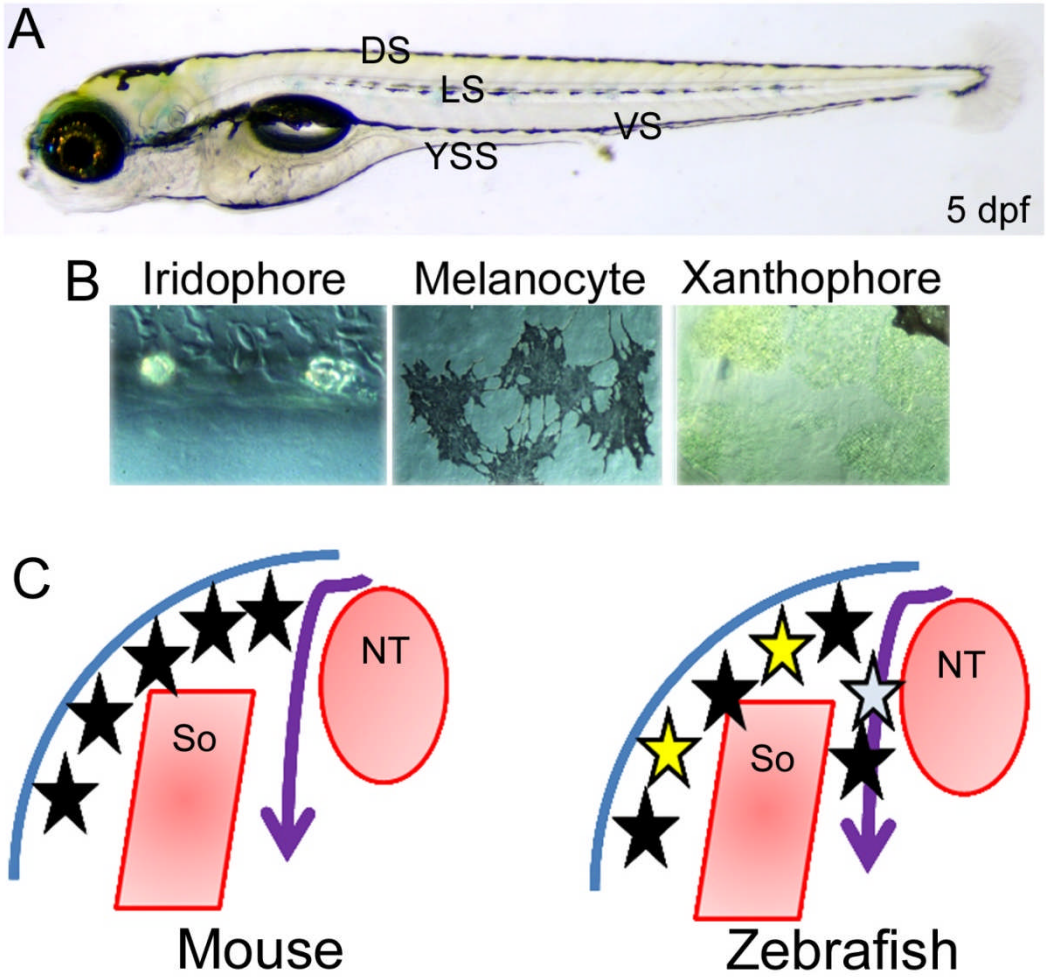


Figure 1: Zebrafish embryonic pigment cell pattern.
(A) Lateral view of a 5 dpf zebrafish embryo with anterior left and dorsal top. Note that melanocytes (black pigmented cells) are visible in four stripes, DS = dorsal stripe, LS = lateral stripe, VS = ventral stripe and YSS = yolk sac stripe. Yellow pigmented xanthophores are also visible particularly in the head and the dorsal region of the trunk. Reflective iridophores are visible in the eye but cannot be seen in this photo. Iridophores are present in the DS, VS and YSS. (B) Images of zebrafish pigment cells. Iridophores are reflective cells containing platelets of guanine, melanocytes are black melanin containing cells and xanthophores are yellow pteridine containing cells. (C) NCCs migrate on stereotypical pathways as shown in schematics of a transverse section through a mouse and a zebrafish trunk. In mice, melanoblasts (black stars) only migrate between the skin and somites (So) on the dorsolateral (lateral) pathway. Neurons and glia of the PNS migrate between the neural tube (NT) and somites (purple arrow) on the ventromedial (medial) pathway. In zebrafish embryos, xanthoblasts (yellow stars) are restricted to the lateral pathway while iridoblasts are restricted to the medial pathway (light blue star). Melanoblasts migrate on both the lateral and medial pathway. Neurons and glia of the PNS migrate on the medial pathway. Figure adapted from Kelsh et al., 2004.

1.1.5 Zebrafish Neural Crest Mutants

A large number of zebrafish NC mutants have been generated, mainly through large scale chemical mutagenesis and insertion mutant screens (Henion et al., 1996, Haffter et al., 1996, Driever et al., 1996, Amsterdam et al., 2004). Many pigment mutants were identified during these screens. Not only are pigment cells easily visualised but the stereotypical embryonic pigment pattern of zebrafish permits differences in cell number, distribution, differentiation and specification in mutant embryos to be identified. The zebrafish WT pigment pattern is comprised of three pigment cell types, melanocytes (black), xanthophores (yellow) and iridophores (reflective or shiny). Melanocytes are organised into four stripes, the dorsal stripe, lateral stripe, ventral stripe and yolk-sac stripe. It should be noted that the melanised cells of the eye in the retinal pigment epithelium (RPE) are not NC derived. Iridophores co-localise with melanocytes in the dorsal, ventral and yolk-sac stripes and migrate to cover the eye. Xanthophores fill in the gaps between the pigment cell stripes but are more concentrated dorsally (**Figure 1**) (Kelsh, 2004). Pigment cell mutants were sorted into several classes as follows (Kelsh et al., 1996): Some mutants, for example *colourless/sox10*, displayed a lack of all pigment cell types (Class 1). In addition *sox10* mutants showed defects in all non-skeletogenic NC derivates including neurons and glia and as such this mutant is of particular interest. Class 2 mutants showed a strong reduction or absence in only one pigment cell type, for example *pffefer* (xanthophores) and *shady/ltk* (iridophores). Additional mutants that only show defects in one pigment cell type have also been identified, for example *nacre/mitfa* (melanocytes) (Lister et al., 1999). Some mutants (Class 3) displayed reduced melanocytes, such as *sparse/kita*, although during the Kelsh et al., 1996 screen no mutants lacking melanocytes were identified. Defects in five genes led to pigment pattern defects (Class 4), this included *choker*. The fifth class of mutants (Class 5) showed additional pigment cells in ectopic locations, for example *parade*. The largest class of mutants discovered (Class 6) displayed defects in pigment cell pigmentation and was split into several subclasses. Examples of genes placed into this category are *golden/slc24a5, albino* and *touch-down*. It should be noted that a large number of mutations affecting xanthophores were identified (Kelsh et al., 1996, Odenthal et al., 1996). The location of the lesion of many of these mutants have yet to be identified, thus the defective gene is unknown.

Non-skeletogenic NC derivatives include neurons and glia of the peripheral nervous system (PNS). These cells contribute to the enteric nervous system and sensory neurons and glia of the dorsal root ganglia (DRGs). Mutants with defects in the PNS have been identified. As already mentioned, the *colourless/sox10* mutant displays a loss of all pigment cell types and defects in the PNS (Dutton et al., 2001a). The $sox10^{baz1}$ mutant allele is unique amongst the *sox10* mutant alleles as it has supernumerary DRGs instead

of a reduced number (Carney et al., 2006). The *mother superior/foxd3* mutant displays a depletion of cells in the PNS, pigment cells and craniofacial cartilage (Montero-Balaguer et al., 2006). The *nosedive* mutant shows a lack of DRGs but little effect on other NC derivatives although some melanocytes do appear in ectopic locations where DRGs normally lie (Henion et al., 1996). PNS defects are harder to identify as molecular markers are required to label these NC derivatives. A large number of zebrafish mutants are available, particularly as a result of the large scale mutagenesis screens. These mutants provide a valuable resource to dissect the processes that occur during NC development.

1.1.6 Waardenburg syndromes and Hirschprung's disease

Defective NC development leads to a number of syndromes collectively known as neurocristopathies. The Waardenburg syndromes are a major family of neurocristopathies and patients display defects in a number of NC derivatives. The symptoms of these defects include varying combinations of hypopigmentation and sensorineural hearing loss. Sensorineural deafness results from defects in the inner ear, specifically the cochlea. The Waardenburg syndromes are split into four sub-types based on the combination of phenotypes presented. A number of mutations that are causative of these diseases have been located, all of which identify key factors during neural crest development.

1.1.6.1 Waardenburg syndrome Type I (WS1)

WS1 (OMIM #193500) patients display pigment abnormalities such as partial hair albinism and heterochromatic irises, subtle facial defects such as dystopia canthorum (inner canthi of the eyes are displaced further apart than normal), a broadening of the nose and sensorineural deafness. WS1 shows autosomal dominant inheritance and although dystopia canthorum is the most consistent feature of this syndrome, the symptoms presented and severity of the disorder vary from case to case. WS1 is caused by lesions in the *PAX3* gene, thus the mouse *Pax3/Splotch* mutant is a model of this disease (Tassabehji et al., 1992). *Splotch* heterozygous mice show pigment abnormalities but not deafness while *Splotch* homozygous mice show very severe defects and die prior to birth (Tachibana et al., 2003).

1.1.6.2 Waardenburg syndrome type II (WS2)

WS2 was initially described to classify Waardenburg syndrome cases that did not display the characteristic phenotype of WS1, dystopia canthorum. Cases of WS2 typically only show eye and skin pigment abnormalities and deafness. WS2 has now been split into a number of sub-types: WS2A (OMIM #193510), WS2B (OMIM %600193), WS2C (OMIM

%606662), WS2D (OMIM #608890) and WS2E (OMIM #611584). The loci affected in WS2B and WS2C have been mapped but no gene has yet been associated with these sub-types. The very rare form of WS2, WS2D, is caused by deletions in the *SNAI2* gene. The original WS2, WS2A, is caused by lesions in the melanocyte master regulator gene *MITF* (Tassabehji et al., 1994). WS2A follows a pattern of autosomal dominant inheritance and is caused by a haploinsufficiency of MITF (Nobukuni et al., 1996). Both the mouse mutant *Mitf/mi* and the zebrafish mutant *mitfa/nacre* are animal models of this disorder. It should be noted that the zebrafish *nacre* mutant results from mutations in both copies of the *mitfa* gene and not from haploinsufficiency. The pigment and hearing defects associated with WS2A can both be explained by a lack of melanocytes. In regard to hearing, a loss of melanocytes in the cochlea can result in deafness (Nobukuni et al., 1996). A recently identified form of WS2 is WS2E and results from a mutation in *SOX10* (Bondurand et al., 2007).

1.1.6.3 Waardenburg syndrome type III (WS3; Klein-Waardenburg Syndrome)

As with WS1, WS3 (OMIM #148820) is caused by mutations at the *PAX3* locus. WS3 shares the same defects as observed with cases of WS1 but in addition patients present with musculoskeletal defects that result in limb abnormalities. Thus WS3 patients present with a more severe phenotype (Hoth et al., 1993). There is evidence that WS3 occurs when two defective copies of *PAX3* are inherited. Homozygous *Splotch* mice may represent a model for WS3 and these mice do display limb abnormalities along with severe neural tube defects which are absent in humans (Zlotogora et al., 1995).

1.1.6.4 Waardenburg syndrome type IV (WS4; Waardenburg-Shah syndrome)

WS4 (OMIM #277580) combines the pigmentation defects of WS2 with the enteric aganglionosis of Hirschsprung disease (OMIM #142628). Hirschsprung disease patients show a congenital absence of ganglion cells along the gut, particularly in the terminal portion of the gut. Thus multiple NC derivatives are affected in this neurocristopathy. WS4 can be inherited in either a recessive or dominant manner depending on the causative mutation. The recessive condition typically results from mutations in the G-protein coupled receptor *EDNRB* or the ligand *EDN3* (Edery et al., 1996, Hofstra et al., 1996, Syrris et al., 1999). The recessive WS4 condition presents with a range of severities but always with some degree of hypopigmentation and often enteric aganglionosis. Sensorineural deafness is seen in recessive WS4 patients but tends to be rare and may be caused by a lack of melanocytes in the cochlea (Hofstra et al., 1996). The mouse mutants *Ednrb/piebald-lethal* and *Edn3/lethal spotting* are models for WS4 (Tachibana et al., 2003). This disorder and the phenotype of the mouse mutants suggests at a conserved

role for *EDNRB/EDN3* signalling in both melanocyte and enteric nervous system development. The dominant form of WS4 is caused by heterozygosity at the *SOX10* locus resulting in haploinsufficiency (Pingault et al., 1998). This leads to hypopigmentation combined with Hirschsprung disease and is often associated with deafness. Deafness in Waardenburg syndromes is usually associated with a loss of melanocytes in the cochlea but *SOX10* is expressed in the ear thus deafness may result from a direct effect (Bondurand et al., 1998). The mouse mutant *Sox10/Dominant megacolon* and the zebrafish mutant *sox10/colourless* provide models of WS4 and will be examined in more detail later.

1.1.6.5 Waardenburg-Shah syndrome, neurologic variant

A neurologic variant of Waardenburg-Shah syndrome, also called peripheral demyelinating neuropathy, central dysmyelinating leukodystrophy, Waardenburg syndrome and Hirschsprung disease (PCWH), has been characterised (OMIM #609136) (Inoue et al., 2004). This disease combines the following four disorders that display defects in characteristic cell types; peripheral demyelinating neuropathy (Schwann cells), central dysmyelinating leukodystrophy (oligodendrocytes), Waardenburg syndrome (melanocytes) and Hirschsprung disease (enteric ganglia). PCWH combines the defects of WS4 patients with myelin deficiency (Inoue et al., 2002). PCWH is caused by mutations in the *SOX10* gene (Inoue et al., 2002, Inoue et al., 2004). Mutations resulting in a truncated SOX10 protein can cause WS4 or the more severe PCWH. In WS4 cases the truncated *SOX10* mRNA was degraded, while in PCWH cases truncated mRNA showed more stability and was translated into a protein with dominant-negative activity (Inoue et al., 2004).

1.2 Transcription Factors

1.2.1 What are transcription factors and how do they function?

Eukaryotic RNA polymerase II cannot initiate transcription without additional factors. General transcription factors bind to a gene's promoter sequence to facilitate the tight binding of RNA polymerase thus allowing transcription to commence (Alberts et al., 1998). Therefore transcription factors (TFs) are DNA binding proteins that help to initiate the transcription of a gene. These TFs recognise common promoter sequences and are present in all cells thus do not display any specificity regarding the genes that they activate. To achieve specificity in the activation of transcription, genes contain regulatory sequences within their promoters that are recognised by specific transcription factors (Alberts et al., 1998). Therefore a second key feature of TFs is the ability to recognise and

bind tightly to a specific DNA sequence. These TFs regulate a subset of genes as a consequence of a particular cellular environment. In this way a transcription factor (TF) can regulate the expression of genes required for a unique purpose such as directing a cell to specify and differentiate into a particular cell type. It should be noted that some TFs activate gene expression while others act as repressors, indeed some transcription factors can perform both roles under different regulatory conditions (Latchman, 1997).

TFs all contain a DNA binding motif but can be separated into different groups based on sequence homology and tertiary protein structure. Examples of these include helix-loop-helix TFs such as bHLH genes, leucine zipper TFs, zinc finger TFs and helix-turn-helix TFs such as *Hox*, *Pax* and *Fox* genes. The majority of TF DNA binding motifs insert into the major groove of the DNA double helix and form tight associations with a number of DNA base pairs (Alberts et al., 1998). To regulate transcription, TFs can function through several mechanisms. TFs can stabilize or disrupt the binding of RNA polymerase to the gene promoter to increase or inhibit transcription. TFs can regulate the accessibility of a promoter region to RNA polymerase by altering the conformation of chromatin. For example, TFs can help to regulate histone acetylation. Histone acetylation by histone acetyltransferase can enhance the accessibility of DNA to protein complexes like RNA polymerase while hypoacetylation by histone deacetylase strengthens the association of DNA with histones making it less accessible (Narlikar et al., 2002). TFs can also recruit additional activating or repressing proteins to a promoter mediated through a protein interacting domain (Latchman, 1997). Indeed many TFs work in concert with a number of other TFs and co-activating or co-repressing proteins. These proteins all respond to environmental and biological stimuli to ensure that the transcription of a gene proceeds at the correct rate.

1.2.2 Transcription Factors in the Neural Crest

A network of transcription factors act to induce and specify NCCs during early NC development. This includes NC specifier genes such as *Foxd3* and *Sox10*. Subsequently, a number of TFs that drive the specification and differentiation of specific cell types are activated. Such TFs with roles in NC specification and differentiation, termed master switch TFs, which are currently known, will be examined here.

1.2.2.1 Mitf is required for melanogenesis

The mammalian melanocyte master regulator TF is a basic helix-loop-helix leucine zipper (bHLH-Zip) TF known as *Mitf*. The mouse *Mitf/microphthalmia* mutant displays severe defects in the NC derived melanocyte lineage (Opdecamp et al., 1997). The zebrafish

homologue of *Mitf* is *mitfa* and the *mitfa/nacre* mutant also has an absence of melanocytes (Lister et al., 1999). Melanocytes fail to specify correctly in *mitfa* mutants as indicated by the failure of early melanoblast marker expression. In addition, expression of WT *mitfa* in mutant embryos can rescue melanocyte development (Lister et al., 1999). *Mitf* initiates the melanocyte developmental programme by activating differentiation and survival genes such as *Dct, c-Kit, Ednrb, Tyr* and *Pmel17* (Thomas and Erickson, 2008). Expression of *Mitf* is mainly driven by the NC specifier TFs *Sox10* and *Pax3* (Bondurand et al., 2000). WNT signalling also plays a key role in driving the differentiation of melanoblasts and can up-regulate *Mitf* expression (Thomas and Erickson, 2008).

1.2.2.2 Transcription factors involved in xanthophore development

In mice, *Pax3* plays a role in melanocyte development by activating expression of *Mitf* in concert with *Sox10*. *Pax3* is a paired box transcription factor known to be expressed in very early NCCs and expression is maintained in pre-migratory NCCs (Lewis et al., 2004, Minchin and Hughes, 2008). A recent study has identified a key role for *pax3* in zebrafish xanthophore development (Minchin and Hughes, 2008). When *pax3* expression was knocked down using morpholinos, xanthoblast marker gene expression was lost, suggesting that *pax3* plays a role in xanthophore specification. Given that xanthophore specification also fails in *sox10* mutant embryos (Dutton et al., 2001a) and PAX3 and SOX10 co-operate during mouse melanocyte development (Potterf et al., 2000), it is plausible that these two genes also function together in xanthophores. Expression of early markers of the NC such as *foxd3* and *sox10* were not affected by *pax3* knockdown although at later stages *sox10* expression was reduced in pre-migratory and migrating NCCs in morphant embryos. It was identified that *pax3* was required for *sox10* expression in developing enteric neurons and that *pax3* deficiency resulted in a failure of enteric neuron formation. A role for *pax7* in xanthophore differentiation, but not specification, was also identified. Morpholino knock down of *pax7* did not affect the expression of early xanthoblast markers, such as *xdh*. It was identified that Pax7 did not play a role in xanthophore specification as these cells could be still be detected but they failed to pigment. Thus multiple TFs are required during xanthophore development (Minchin and Hughes, 2008).

1.2.2.3 Transcription factors involved in peripheral nervous system development

The peripheral nervous system (PNS) consists of NC derivatives and a number of transcription factors key to PNS development have been identified. The autonomic nervous system (sympathetic, parasympathetic and enteric ganglia) fails to form in the mouse mutant of the paired-homeodomain TF *Phox2b* (Pattyn et al., 1999). In zebrafish,

knockdown of *phox2b* expression using morpholinos resulted in a strong reduction of enteric ganglia thus showing a conserved role in ENS development between the two organisms (Elworthy et al., 2005). Knockout of the TF *Phox2a* in mice also leads to some autonomic defects but these are not as pronounced as in *Phox2b* mutant mice (Morin et al., 1997, Pattyn et al., 1999). The bHLH TF *Mash1* is also important during autonomic nervous system development. *Mash1* knockout mice display extensive defects in the autonomic nervous system (Guillemot et al., 1993). An *in vitro* study identified a role for *Mash1* in inducing the expression of neuronal differentiation genes following the specification of a neuronal precursor (Sommer et al., 1995). It is thought that *Phox2b* acts upstream of *Mash1* as *Phox2b* knockout mice do not express *Mash1* in the autonomic nervous system (Pattyn et al., 1999).

Sensory neurons, whose cell bodies lie in the dorsal root ganglia (DRGs), are also derived from the NC. The two neurogenin genes, *Ngn1* and *Ngn2*, are proneural bHLH TFs. Neurogenins are expressed in the precursors of sensory neurons although *Ngn1* and *Ngn2* show slightly different temporal and spatial expression patterns. Analysis of mutant embryos identified functional redundancy between these two TFs and only in a double mutant were all sensory neurons absent (Ma et al., 1999). Forced expression of these neurogenins in chick NC biased the cells towards a sensory neuron fate *in vivo* (Perez et al., 1999). This suggested that neurogenins play a role in specifying sensory neurons. Similarly, *ngn1* has been shown to be essential for sensory neuron specification in zebrafish (Carney et al., 2006). Thus a number of TFs play roles in the specification and differentiation of elements of the PNS.

1.3 Introduction to Sox10

1.3.1 The Sox Family Overview

Sox proteins comprise a family of approximately 20 transcription factors with a high-mobility-group (HMG) DNA binding domain (Schepers et al., 2002). The HMG domain was first identified in *SRY*, the sex determining gene on the Y chromosome and *Sox* genes are named to reflect this (*SRY*-related high-mobility-group box). *Sox* genes have been classified into eight groups (A-H) based on sequence homologies both inside and outside of the HMG domain. The HMG domain is almost identical across *Sox* genes within the same group but only approximately 50 % similar between groups (Lefebvre et al., 2007). Sox TFs display sequence specific DNA binding but, unlike most TFs, they bind the minor groove of the double helix. In addition Sox proteins bend DNA and interact with a range of protein partners. Via these mechanisms, Sox TFs play a critical role in the regulation of transcription. *Sox* genes are expressed in a wide range of tissues but also show tightly

regulated temporal and spatial expression patterns. This family of genes is known to regulate multipotency, specification and differentiation in a range of cell types and play key roles during development (Lefebvre et al., 2007).

The HMG box fulfils the functions of DNA binding, DNA bending and in some *Sox* groups, protein interactions and nuclear import or export. As mentioned previously, unlike most TFs, Sox genes interact with the double helix minor groove. All Sox TFs tested display an *in vitro* binding preference for the sequence 5'-$^{A}/_{T}^{A}/_{T}$CAA$^{A}/_{T}$G-3' although different Sox proteins display varying preferences for nucleotides flanking this core sequence (Mertin et al., 1999). It has been shown that Sox proteins can bind *in vivo*, sequences that are only a partial match for this *in vitro* consensus sequence (Lefebvre et al., 1997). This complicates the search for Sox binding sites through *in silico* exploration of genomic data. Thus experimental approaches may be required to identify regulatory targets of *Sox* genes.

Sox proteins can influence target transcription by bending the target promoter sequence. This manipulates the three dimensional structure of the promoter. It is thought that this brings together TFs bound to disparate regulatory sites and other proteins bound to the promoter to alter transcription rates. This may occur by altering the arrangement and interactions of the components of the enhanceosome (functionally active complexes of TFs bound to gene enhancer sequences) (Lefebvre et al., 2007). While solid evidence of this process is required, some mutant studies have provided evidence of the important role that DNA bending plays. A mutation in the *Sox2* HMG box that reduced the bend imparted to the DNA by the TF actually increased target transcription in comparison to WT protein. In addition, mutation of the target enhancer sequence to render the DNA more difficult to bend decreased transcription rates (Scaffidi and Bianchi, 2001). Therefore the architecture of the promoter and the enhanceosome plays an important role in regulating transcription rates.

Sox proteins not only bind to a specific DNA sequence but can also bind to and interact with other proteins and TFs (Wilson and Koopman, 2002). The TFs Sox2 and Oct3 are required to activate transcription of *Fgf4* during early embryonic development (Yuan et al., 1995). An enhancer on the *Fgf4* promoter contains adjacent Sox and Oct TF binding sites. These two TFs act in concert to activate the enhancer. These two TFs actually synergise when bound to DNA through their DNA binding domains to form a heterodimer (Reményi et al., 2003). All interactions between Sox proteins and other TFs studied thus far occur through the DNA binding domain. These interactions, while bound to DNA, may be facilitated by the fact that Sox proteins bind to the minor groove of DNA while most other TFs bind to the major groove. Thus cell specific gene activation by Sox proteins often requires additional partners; this may help to increase specificity despite the

generality of the Sox DNA recognition sequence. Sox TFs can regulate transcription rates both by bending DNA and by regulating the formation of transcriptionally active protein complexes.

In addition to the HMG DNA binding domain, some groups of Sox proteins also include protein interacting domains known as transactivating or transrepression domains in their C terminal region (**Figure 2**). For example, during chrondrogenesis, Sox9 binds to the transcriptional co-activator CBP/p300 to enhance transcription from the *Col2a1* promoter (Tsuda et al., 2003). Transrepression domains have been identified in members of the SoxB group (Uchikawa et al., 1999). Thus Sox proteins are capable of regulating the activity of enhanceosomes and therefore transcription through architectural alterations and direct protein interactions (Lefebvre et al., 2007).

1.3.1.1 The SoxE group

The *SoxE* group is made up of three *Sox* genes, *Sox8, Sox9* and *Sox10* (in zebrafish additionally there is *sox9b*). All three genes are expressed in multiple cell types during embryonic development. During mouse development, *Sox8* is expressed in a number of developing sensory organs including the otic vesicle, the brain, the spinal cord and a number of NC derivatives including cranial ganglia, dorsal root ganglia, branchial arches and the enteric nervous system (Schepers et al., 2000, Sock et al., 2001). A closer examination of expression in the brain identified that *Sox8* marks developing glia (Cheng et al., 2001). Despite widespread expression of *Sox8* and the strong phenotypes displayed by *Sox9* and *Sox10* mutants, a homozygous mutant *Sox8* mouse was still viable, did not display specific defects in *Sox8* expressing tissues and only exhibited reduced weight (Sock et al., 2001). As *Sox8* expression overlaps with expression of other *SoxE* genes this weak phenotype may result, to some extent, from functional redundancy (Stolt et al., 2004). Certainly, *Sox8* and *Sox9* display functional redundancy during testis cord differentiation with *Sox9* knockout mice not displaying a phenotype in this tissue but Sox9 and *Sox8* double mutants displayed a more severe phenotype than *Sox8* mutants (Barrionuevo and Scherer, 2009). *Sox8* is known to function during gliogenesis, male gonad development, osteogenesis and neural crest formation in *Xenopus* (Chaboissier et al., 2004, Stolt et al., 2004, Schmidt et al., 2005, O'Donnell et al., 2006). The human disorder ATR-16 (OMIM #141750) has been associated with deletions of a chromosomal region that includes the area that *SOX8* maps to (Pfeifer et al., 2000). No specific human disease has been attributed to *SOX8* thus this gene has been less well studied than the other two members of this group.

Haploinsufficiency of *SOX9* results in the disorder Campomelic Dysplasia (CMPD, OMIM #114290) which is characterised by skeletal defects, particularly of the long bones,

and many XY individuals also suffer genital defects or develop as females (Wagner et al., 1994). Consistent with these defects, sites of expression of *Sox9* in mouse embryos strongly correlate with sites of chondrogenesis and chondrocyte precursor cells. For example, at twelve days post coitum most skeletal structures show *Sox9* expression (Wright et al., 1995). *Sox9* was identified as an essential transcription factor in chondrocyte specification and differentiation and therefore in the development of cartilage and skeletal structures (Bi et al., 1999). Also consistent with the phenotype of CMPD, *SOX9* is thought to play a role in sex determination. *Sox9* has been shown to be essential for Sertoli cell differentiation and activates the expression of several Sertoli cell specific markers (Chaboissier et al., 2004). *Sox9* is also expressed in the early NC and forms part of the NC specifier gene network. Forced expression of *Sox9* in chick embryos can induce the expression of NC markers and the formation of trunk NC neural derivatives fails in *Sox9* knockout mice. In these mice, NCCs were shown to apoptose after delamination (Cheung et al., 2005).

The third member of the *SoxE* family is *Sox10* and is associated with the human disorders WS4 and PCWH. Sox10 contains an HMG DNA binding domain and at the C terminus a transactivation domain (**Figure 2)**. *Sox10* has been termed a NC specifier gene, playing a critical role in the specification of all non-skeletogenic NC derivatives (Dutton et al., 2001a). *Sox10* also functions during cellular differentiation and in maintaining cell multipotentiality, although the later function has yet to be substantiated in zebrafish (Britsch et al., 2001, Kim et al., 2003, Carney et al., 2006). In addition, *sox10* plays a role in inner ear formation (Whitfield et al., 1996). This transcription factor will be examined in more detail in the next few sections. The *SoxE* group of TFs regulate a large number of important developmental processes and have been associated with a number of human developmental disorders. Therefore this group of genes have been and remain of tremendous interest.

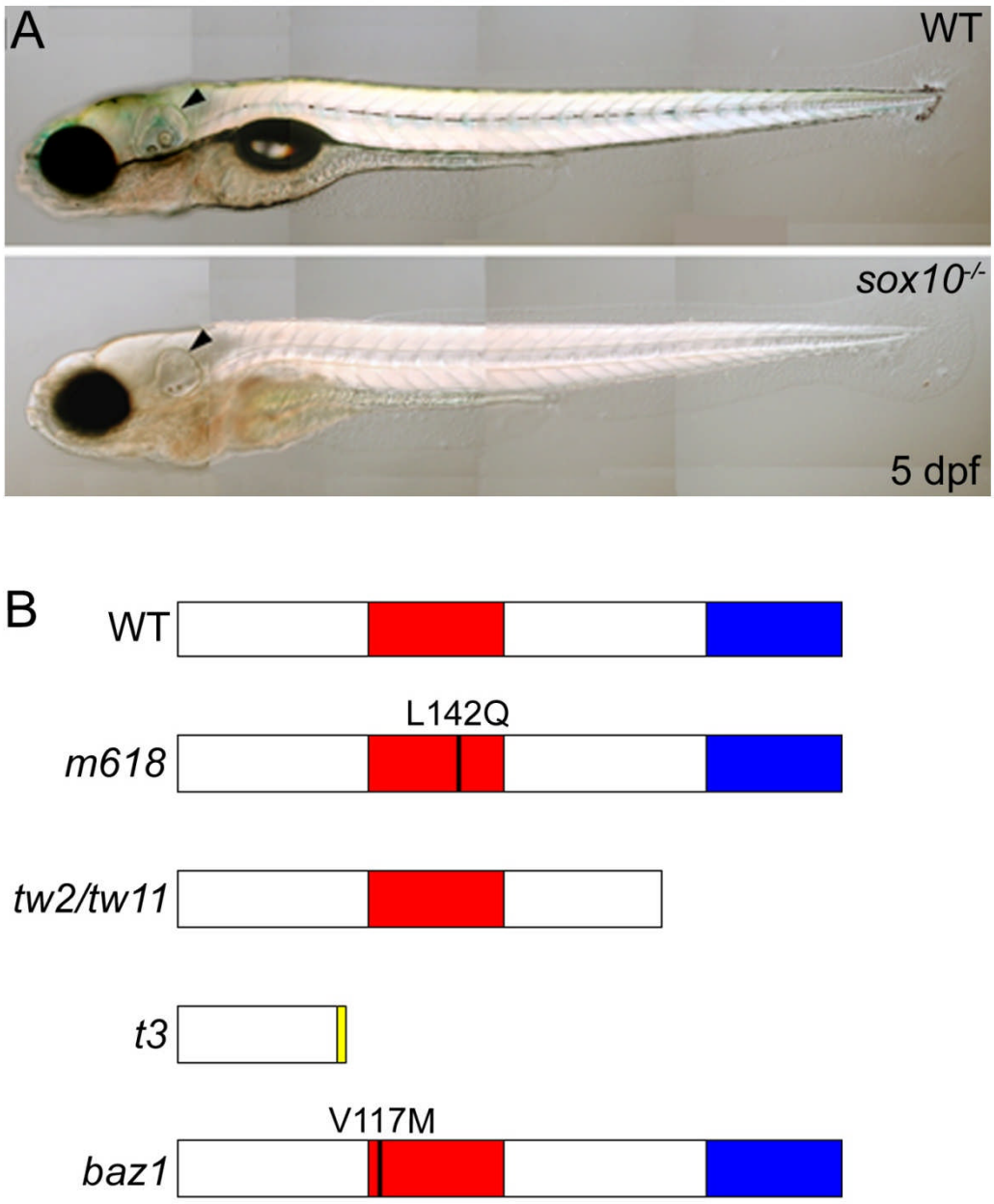


Figure 2: Zebrafish sox10 mutant.
(A) Lateral image of a 5dpf zebrafish WT embryo (top) and *sox10* mutant (bottom) embryo. Note the lack of pigmented cells in the mutant embryos, particularly striking is the lack of dark melanocytes in the head and trunk. The eye is still pigmented by melanocytes as these are not NC derived and are thus not Sox10 dependent. The otic vesicle in both embryos is marked by an arrowhead; note the small otoliths in the mutant embryo. (B) Schematics of WT Sox10 protein and Sox10 mutant proteins produced by the different *sox10* mutant alleles. Red box = HMG domain, blue box = transactivation domain. The yellow box of the t3 allele protein denotes additional amino acids inserted as part of the frame shift mutation prior to protein termination. The positions of the leucine to glutamine and valine to methionine substitutions in the m618 allele and baz1 allele proteins respectively are shown. The tw2 and tw11 alleles both produce truncated proteins lacking the transactivation domain.

1.3.2 Expression of *Sox10*

The expression pattern for *Sox10* has been determined for a number of vertebrate species by *in situ* hybridization and Northern blotting and appears to be conserved across species. In humans, *SOX10* is expressed in the CNS of adults. During embryonic development *SOX10* expression was noted in the otic vesicle and trunk NCC derivatives. While expression in melanocytes was not directly observed, strong expression in the PNS was seen (Bondurand et al., 1998). *Sox10* expression has been examined further in a range of model organisms including rodents, chick, *Xenopus* and zebrafish (Southard-Smith et al., 1998, Kuhlbrodt et al., 1998, Cheng et al., 2000, Dutton et al., 2001a, Aoki et al., 2003). The sites of *Sox10* expression outside of the NC in the otic vesicle and oligodendrocytes, are conserved across all organisms. *Sox10* expression is particularly strong in the otic vesicle. In the NC *Sox10* again displays a conserved expression pattern across species, although these expression patterns have not been directly compared. In all organisms *Sox10* is expressed in pre-migratory NCCs and it seems likely that *Sox10* is a generic marker of non-skeletogenic NC. Expression has also been seen in migrating NCCs on the medial and dorsolateral pathways in the trunk and in ENS precursor cells migrating along the developing gut. It has been noted that *Sox10* expression is down-regulated in differentiating NCCs, for example *sox10* expression is lost in zebrafish as melanocytes pigment (Dutton et al., 2001a, Greenhill 2008). In contrast, *Sox10* expression persists in differentiating glia even until adulthood (Kuhlbrodt et al., 1998). Thus *Sox10* is expressed in a range of NC non-skeletogenic derivatives and non-NC derivatives and shows dynamic changes in expression as development proceeds.

The expression of *sox10* in zebrafish (**Figure 3**) has been examined in detail (Dutton et al., 2001a). Cells expressing *sox10* are first visible at approximately 10 hpf at the lateral edge of the neural plate. By 11 hpf cranial NCCs are marked by *sox10* expression and as development progresses, *sox10* expression spreads caudally and is seen in trunk NCCs at around 16 hpf. Up until this stage all *sox10* expressing cells in the trunk have yet to begin migrating. The first NCCs to migrate are neural derivatives on the medial pathway at approximately 16 hpf (Raible et al., 1992). These cells initially express *sox10* during migration. At 24 hpf *sox10* expression can be seen in pre-migratory NC and NCCs migrating on the medial pathway, in the otic vesicle, in two clusters in the head that correspond to cranial ganglia and in cells along the posterior lateral line nerve (**Figure 3**). Scattered cells in the head express *sox10* in a pattern that bears a striking resemblance to markers of pigment cells. Only rarely are *sox10* expressing cells seen on the lateral pathway, perhaps as a consequence of expression being rapidly down-regulated in pigment cells. If *in situ* hybridizations are pushed hard, these cells can be visualised. Precursor cells of the ENS can be seen expressing *sox10* at 36 hpf and Sox10 positive cells are seen forming the DRGs by 40 hpf (Elworthy et al., 2005, Carney et al., 2006).

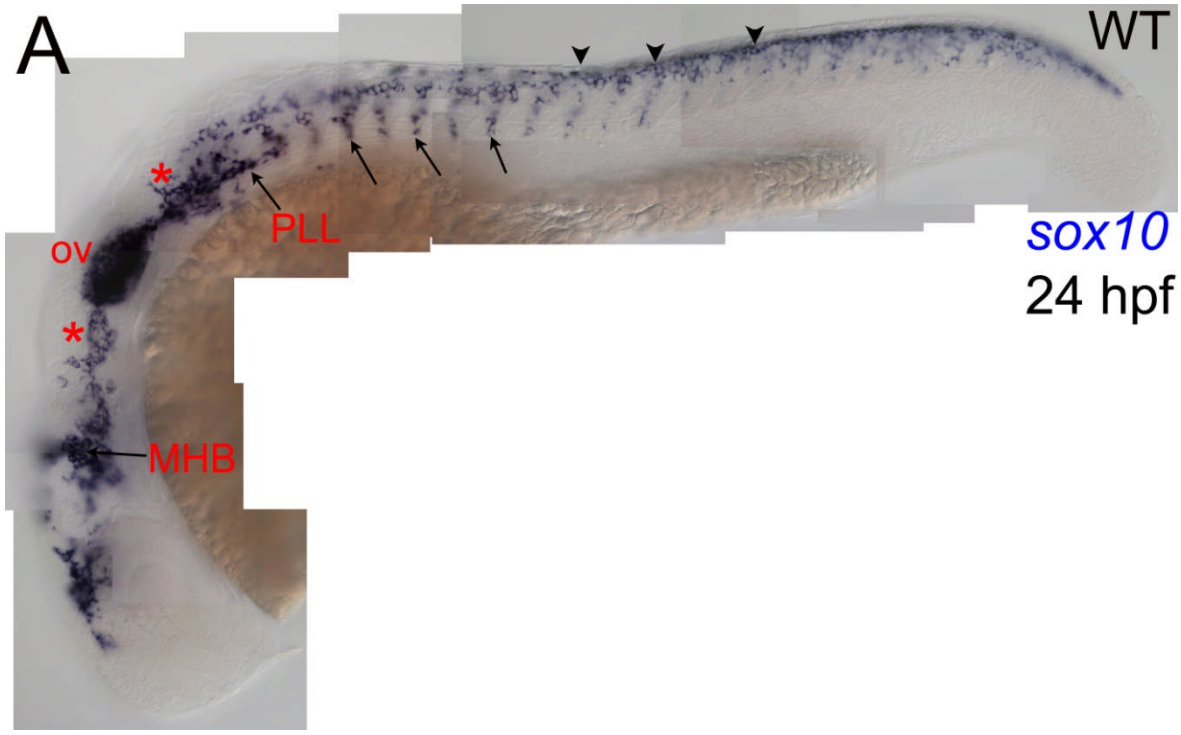


Figure 3: Whole mount *in situ* hybridization of *Sox10* expression.
Lateral view of a 24 hpf zebrafish embryos is shown, anterior to left and dorsal top. Expression in NCCs migrating at the midbrain-hindbrain boundary (MHB), otic vesicle (ov) and glia of the posterior lateral line nerve (PLL) is marked. In addition expression in two patches either side of the otic vesicle are marked by asterisks, cells at these locations form the cranial ganglia. Arrowheads mark pre-migratory NCCs in the trunk and arrows mark NCCs migrating on the medial pathway.

1.3.3 *Sox10* mutants

The spontaneous mouse mutant *Dominant megacolon* (*Dom*) was mapped to the *Sox10* locus. As a result of a frame shift mutation in which an extra guanine was inserted C terminal of the HMG box domain, 99 novel amino acids were added to the protein before terminating (Herbarth et al., 1998, Southard-Smith et al., 1998). Functional analysis of this mutant protein showed that DNA binding was not affected but transcriptional activation of partner proteins mediated by the transactivation domain was abolished (Herbarth et al., 1998). Heterozygotes of this mutant display pigment abnormalities and megacolon, homozygotes are embryonic lethal. In addition, heterozygous animals display behavioural defects associated with deafness (Pingault et al., 1998). It is thought that deafness in WS4 and the *Dom* mouse results from a loss of melanocytes in the ear and is a secondary consequence of a failure in *Sox10* signalling. Analysis of NC markers identified defects in a number of NC derivatives in *Dom* embryos with homozygotes more severely affected than heterozygotes. *Dom/Dom* embryos lack Dct^{+} melanoblasts and also display a lack of *Ednrb* and *Sox10* expressing NCCs. These expression patterns identified defects in the PNS, particularly a failure of enteric neurons to colonise the gut alongside pigment abnormalities. Cranial ganglia were disrupted and displayed morphological defects in *Dom/Dom* embryos but this was not observed in heterozygous embryos (Herbarth et al., 1998, Southard-Smith et al., 1998). Significant levels of NCC apoptosis were identified in both heterozygous and homozygous *Dom* mutant embryos (Southard-Smith et al., 1998).

A targeted *Sox10* mouse mutant has also been constructed by knock in of the *lacZ* reporter gene that deleted the entire *Sox10* reading frame (Britsch et al., 2001). Heterozygous and homozygous $Sox10^{lacZ}$ mice display essentially the same phenotypes as heterozygous and homozygous *Dom* mice respectively. This indicated that the *Sox10* mutant phenotype could be caused by haploinsufficiency and not necessarily as a result of possible dominant negative activity of the *Dom* mutant SOX10 protein. In addition, *LacZ* staining was seen in the same sites as *Sox10* expression as detected by *in situ* hybridization. This included sites outside the NC such as in oligodendrocyte precursors in the CNS and the otic vesicle and NC derived cell types in the PNS and melanoblasts. The generation of this mouse line has facilitated the analysis of *Sox10* function in various cell types. For example, analysis of melanocyte development using the melanoblast markers *Mitf, c-Kit* and *Dct* identified a 50 % decrease of melanoblasts in heterozygous embryos and an almost total absence of melanoblasts in homozygous mutant embryos (Britsch et al., 2001). Analysis of glial differentiation markers in $Sox10^{lacZ}$ embryos indicated that cells expressing the reporter gene in DRGs were glial cells and this cell type was reduced in homozygous mutant embryos. Within these DRGs, the sensory neurons were seen to form and then degenerate. Interestingly the authors concluded that the sensory neuron phenotype of *Sox10* mutants was an indirect effect resulting from the lack of glial support

cells (Britsch et al., 2001). The targeted *Sox10* mutant mouse was also used to study the role of SOX10 during ENS development (Paratore et al., 2002). This study identified that enteric aganglionosis resulted from ENS progenitor cells failing to correctly specify in homozygous *Sox10* mutant animals, with the defect in NCC migration a secondary effect. Both neurons and glia failed to specify in the ENS, in contrast to sensory neurons. A second role for SOX10 during ENS development was identified by fate mapping of cells expressing β-galactosidase. While these cells migrated to the gut in heterozygous mutant embryos they did not migrate along the gut and failed to express the NC progenitor markers *Sox10* and *ErbB3*. These cells also failed to adopt appropriate ENS fates. Thus these cells seemed unable to maintain a multipotent progenitor state and could not respond correctly to environmental cues present in the gut. This resulted in a reduced number of progenitor cells available to colonise the gut. The mouse *Sox10* mutants indicate that Sox10 may play various roles in different NC cell types including cell specification, driving cell differentiation, cell survival and maintaining multipotency.

The zebrafish *sox10* mutant, also known as *colourless*, was initially identified during the large scale Tübingen mutagenesis screen (Kelsh et al., 1996). Several mutant alleles have been identified, the *tw2* and *tw11* alleles lack the transactivation domain, and the *t3* allele produces a severely truncated protein that lacks both the HMG and transactivation domain. The *m618* allele generates a full length Sox10 protein but a single base pair substitution has resulted in a non-conservative leucine to glutamine amino acid change within the HMG DNA binding domain (**Figure 2**). This renders the protein effectively non-functional, although this has not been explicitly tested (Dutton et al., 2001a). All of these alleles display a striking phenotype (**Figure 2**) that bears a strong resemblance to the mouse homozygous *Dom* mutant. All of these alleles are recessive and homozygous lethal. All NC derived pigment cells are absent in *sox10* mutant embryos and there is a strong reduction in the expression of key melanocyte specification genes (Dutton et al., 2001a). The pigment defects are combined with PNS defects; enteric ganglia are absent in *sox10* mutant embryos while DRGs are reduced in number and often appear in ectopic positions. Glia are also strongly reduced in mutant embryos (Carney et al., 2006). NCCs form in normal numbers, subsequently non-skeletogenic NCCs in zebrafish *sox10* mutant embryos fail to migrate and die by apoptosis. This cell death occurs during a discrete time period that corresponds to cell differentiation in WT embryos. These defects appear to be secondary consequences of NCCs failing to correctly specify in *sox10* mutant embryos, at least in pigment cell precursors (Dutton et al., 2001a). Zebrafish *sox10* mutants also show a defect in otic vesicle development with a small otic vesicle and small otoliths (Whitfield et al., 1996). Key otic vesicle patterning genes display de-regulated expression patterns in *sox10* mutants thus suggesting *sox10* may play a direct role in ear development beyond the specification of melanocytes within

the cochlea (Dutton et al., 2009). Thus the zebrafish and mouse *sox10* mutants both display strong defects in non-skeletogenic cell types.

Recently a new *sox10* mutant allele was identified in zebrafish, the *sox10*^{baz1} allele. This allele has a valine to methionine amino acid substitution in the HMG DNA binding domain (**Figure 2**) (Carney et al., 2006). The pigment phenotype of *sox10*^{baz1} mutant embryos is slightly less severe than for other mutant *sox10* alleles with melanocytes absent but xanthophores and iridophores mildly reduced. Similar to melanocytes, glia were also strongly reduced in the *sox10*^{baz1} mutant. Surprisingly the number of sensory neurons (DRGs) actually increased in mutant embryos in comparison with WT siblings. This strongly suggests that *sox10* plays a direct role in regulating the specification of sensory neurons, in contrast with evidence gathered from mouse *Sox10* mutants. Induction of early sensory neuron markers also fails in zebrafish *sox10* mutants supporting a role for *sox10* in specifying sensory neuron fate (Carney et al., 2006). Thus it appears that *sox10* functions to specify all non-skeletogenic NC derivatives. The function of *sox10* will be evaluated in more detail in the next section.

1.3.4 Sox10 Function

Sox10 has been postulated to function at a number of stages during NC development including formation of the NC, maintenance of multipotency and NCSC characteristics, cell specification and cell differentiation (for a review see Kelsh 2006). Evidence for the role of *Sox10* in all of these processes comes from various experiments in different model species and from examination of *Sox10* mutant animals. In this section the function of *Sox10* in each of these processes will be examined.

1.3.4.1 Sox10 and formation of the NC

Work in frog has indicated that *Sox10* plays a role in early NC formation. Morpholino mediated knockdown of Sox10 function resulted in a strong reduction of early NC marker gene expression (*Snai2*, *Foxd3* and *Sox9*) (Honore et al., 2003). Over expression of *Sox10* in *Xenopus* embryos was also capable of expanding the expression domain of *Snai2* and *Sox9* (Honore et al., 2003, Aoki et al., 2003). Thus Sox10 appears to expand the number of early NCCs. These results are difficult to interpret as the complicated gene network responsible for NC induction and formation has yet to be fully resolved. The genes involved in this network are known to be interdependent thus it is difficult to identify the specific functions of each individual gene. For example, another *SoxE* gene, *Sox9*, is also expressed in early NCCs and over expression of *Sox9* can induce expression of early NC marker genes including *Sox10* (Cheung and Briscoe, 2003). *In situ* hybridization

analysis has also shown that *Sox10* expression initiates after that of other early NC marker genes, such as *Sox9* (Cheung and Briscoe, 2003). It has also been shown during an animal cap assay in which NC marker gene expression was induced using noggin and Wnt5 that the expression of these marker genes was independent of Sox10 function knockdown using morpholinos (Honore et al., 2003). In mouse and zebrafish *Sox10* mutants, NCCs form in normal numbers strongly suggesting that in these vertebrate species *Sox10* is not required for NC formation (Dutton et al., 2001a, Southard-Smith et al., 1998). There could be species specific differences between the function of *Sox10*. Alternatively, species specific variation in the degree of functional redundancy between *SoxE* genes could exist. Thus in mouse and zebrafish the early roles of *SoxE* genes could largely be functionally redundant while in frogs *Sox10* has a more pronounced early role. In conclusion, early NC formation maybe independent of Sox10 function or the function of Sox10 could be redundant with other *SoxE* genes. However, *Sox10* is expressed in early NCCs and forms part of this complicated NC specifier gene network. It has been hypothesised that *Sox10* functions to maintain NC marker gene expression and therefore NC characteristics rather than in NC formation (Kelsh, 2006).

1.3.4.2 Sox10 and Stem Cell maintenance

Sox10 is expressed in a number of embryonic neural crest stem cells (NCSCs) isolated from pre-migratory NC, DRGs and the gut (Paratore et al., 2001, Iwashita et al., 2003, Kim et al., 2003) and adult NCSCs termed epidermal NCSCs (EPI-NCSCs) and skin derived precursors (SKPs) (Sieber-Blum et al., 2004, Wong et al., 2006). NCSCs can differentiate into PNS neurons, glia and smooth muscle in culture (Stemple and Anderson, 1992). The differentiation of these cells can be directed by exposure to appropriate growth signals, for example BMP2 (bone morphogenic protein) signalling drives neuronal fate choice, glial growth factor (GGF) signalling drives glial fate choice and transforming growth factor-β (TGF-β) signalling promotes smooth muscle fate (Shah and Anderson, 1997). In embryonic NCSCs *Sox10* has been shown to play a role in maintaining multipotency (Kim et al., 2003). This study identified a correlation between a loss of multipotency and a loss of *Sox10* expression in response to growth factor signalling. Forced expression of *Sox10* ensured NCSCs in culture retained the ability to differentiate into neurons or glia despite exposure to a growth factor that would normally direct cells towards a specific fate choice (Kim et al., 2003). In other words, cells maintained multi-potentiality, possibly by ensuring the maintenance of growth factor receptor expression.

To promote sympathetic neuronal differentiation *Sox10* induces the expression of the proneural genes *Mash1* and *Phox2b*. These TFs activate downstream differentiation genes such as *Phox2a*. *Sox10* expression is down-regulated in $MASH1^{+}$ $PHOX2B^{+}$ cells

and forced expression of *Sox10* inhibited neuronal differentiation of these cells. This correlated with a repression of the differentiation marker *Phox2a* (Kim et al., 2003). The transient nature of *Sox10* expression in other NCC lines (except glia) suggests this role may be general and recent work has shown *Sox10* has to be down-regulated in melanocytes for differentiation to proceed (Greenhill, 2008). However, in adult melanocyte stem cells *Pax3* has been shown to activate a melanogenic gene cascade while inhibiting differentiation by repressing *Dct* (Lang et al., 2005). The data above has provided strong evidence that, at least in NCSCs with neural potential, *Sox10* maintains multipotency and inhibits differentiation. It would be very interesting if this could be extended to NCSCs that have the potency to form other cell types such as pigment cells. The adult EPI-NCSCs and SKPs which are $Sox10^+$ may present an opportunity to evaluate this. Indeed, the role of *Sox10* in these cell types has not yet been examined.

1.3.4.3 Sox10 and fate specification

A large body of evidence from both mouse and fish points to a general role for *Sox10* in the specification of a subset of NC derivatives. Fate specification requires the activation of a master switch transcription factor that is both necessary and sufficient to drive the development of a specific cell type. The expression of this master switch TF is regulated by both intrinsic cellular factors, such as Sox10, and extrinsic factors such as extracellular growth factors signalling through receptors. Together these factors activate expression of the appropriate master switch TF which activates the cascade of genes necessary for the differentiation, migration and survival of a specific cell type (**Figure 4**).

The fate specification of melanocytes has been well studied in both zebrafish and mouse model systems and in both systems *Sox10* plays a very early role in melanocyte development (Dutton et al., 2001a, Southard-Smith et al., 1998). *Sox10* is also known to directly regulate the melanocyte master switch TF, *Mitf* in mouse (Verastegui et al., 2000) and *mitfa* in zebrafish (Elworthy et al., 2003). *Mitf* has been well characterised as the pivotal gene in melanocyte development (Levy et al., 2006) and in zebrafish *mitfa* can rescue melanocyte development in *sox10* mutant embryos (Elworthy et al., 2003). Expression of *Mitf* or *mitfa* is absent in both mouse and zebrafish *sox10* mutant embryos (Southard-Smith et al., 1998, Kelsh and Eisen, 2000) and the direct regulation of *mitfa* by *sox10* is crucial for the correct expression of this master switch TF (Elworthy et al., 2003). Loss of function morpholino experiments in frog has also indicated that *Sox10* may specify melanocytes but the early role of *Sox10* during NC formation in this model organism has complicated this analysis (Aoki et al., 2003).

While strong evidence exists that *Sox10* acts in the specification of the melanocyte lineage, work has been ongoing to assess if this model can be generalised to other non-

skeletogenic NCC types. While this body of work is currently less extensive, fate specification of sympathetic, enteric and sensory neurons has been shown to be defective in *Sox10* mutants. Sympathetic neuron specification requires the action of two master switch TFs, *Mash1* and *Phox2b* (Lo et al., 1998, Pattyn et al., 1999). Both of these transcription factors require *Sox10 in vivo* for induction (Kim et al., 2003). *Phox2b* also plays a key role in the specification of enteric neurons in both mouse and zebrafish (Elworthy et al., 2005, Pattyn et al., 1999). Morpholino knockdown of Phox2b function led to a loss of enteric neurons, a phenotype also seen in *sox10* mutant embryos. Expression of *phox2b* is defective in *sox10* mutant embryos thus *sox10* activation of *phox2b* is required for enteric neuron specification. The sensory neuron lineage requires the action of the Neurogenin TFs for specification (Ma et al., 1999, Perez et al., 1999). In both mouse and zebrafish *Sox10* mutants, sensory neuron numbers are reduced although this defect is less pronounced than for other NCC derivatives (Southard-Smith et al., 1998, Dutton et al., 2001a, Carney et al., 2006). It was hypothesised from work in mouse that the defects in sensory neurons were a secondary consequence of the loss of their supporting glia. In zebrafish, analysis of *ngn1* expression identified that sensory neuron specification fails in *sox10* mutant embryos. Sox10 can also induce expression of *ngn1 in vivo* (Carney et al., 2006). Interestingly morpholino knock-down of *sox9b* increased the severity of the *sox10* mutant sensory neuron phenotype (Carney et al., 2006). The neurogenic phenotype of the *sox10*^{baz1} mutant that possess additional sensory neurons but lacks glia also argued against the sensory neuron defect simply resulting from a loss of support cells (Carney et al., 2006). Thus in zebrafish at least, it appears that *sox10* functions during sensory neuron specification. To fully confirm the generalisation of the *Sox10* specification model this work needs to be replicated in mouse.

Despite strong evidence supporting a role for *Sox10* in the specification of non-skeletogenic NCCs, the complex phenotype of *Sox10* mutants could be explained by *Sox10* functioning at multiple steps in development. For example, zebrafish *sox10* mutants show defects in cell differentiation, migration and survival. However, if specification is defective, it is likely that defects in these processes would be a secondary consequence of cells failing to initiate appropriate molecular cascades. For example, in chick at least, lateral pathway migration is dependent on melanocyte specification (Santiago and Erickson, 2002). It is likely that migration is dependent on specification activated pro-migration genes such as receptors for migration cues and cell surface proteins that regulate adhesion. The apoptosis of NCCs in *Sox10* mutants is also likely to be a result of a failure in specification. For example, expression of the zebrafish *c-kit* homologue (*kit*) which is essential for melanoblast survival is absent in *sox10* mutant embryos (Parichy et al., 1999, Dutton et al., 2001a). Importantly, NCC death in *sox10* mutant embryos occurs a considerable period of time after *mitfa* expression fails to be

induced thus indicating that the survival defect is a consequence of specification failing (Dutton et al., 2001a). The differentiation defect of melanocytes in *Sox10* mutants is easily explained by the failure of *Mitf* expression. *Mitf* is known to activate the expression of a large number of enzymes involved in melanogenesis (Levy et al., 2006). However, this specification model may not be as clear cut as this. *Sox10* is known to directly regulate a key gene in melanin synthesis, *Dct*, at least in mouse cells *in vitro* (Jiao et al., 2004, Potterf et al., 2001). While *Mitf* plays a key role in activating *Dct* expression, the action of *Sox10* can enhance this (Jiao et al., 2004). However, *in vivo* work in zebrafish suggested that *mitfa* alone is sufficient to drive melanocyte differentiation as *mitfa* rescued melanocytes in *sox10* mutant embryos (Elworthy et al., 2003). The expression of *sox10*, while overlapping with that of *dct* is down-regulated in differentiating melanocytes and Sox10 can repress the expression of *mitfa* target genes in zebrafish (Greenhill, 2008). Thus while *Sox10* may play a transient role in melanocyte differentiation, this is currently unclear. A similar late role for *Sox10* has also been identified during neurogenesis. *Sox10* both activates expression of the pro-neural genes *Mash1* and *Phox2b* while inhibiting expression of their downstream target gene *Phox2a* (Kim et al., 2003). It remains to be established if *Sox10* has a role during differentiation in other NC derivatives.

According to the model of Sox10 functioning in cell specification, NCC death in *Sox10* embryos occurs as a result of specification failing. Thus cells cannot develop correctly and will not express the appropriate survival factors and subsequently apoptose. It is unclear what the molecular characteristics of these unspecified cells will be but it is possible they will maintain the expression of early NCC characteristics. These genes will fail to be repressed without functional Sox10 and without development proceeding correctly. Thus such unspecified cells may express markers of early NCCs, NCSC genes, receptors for instructive signals and markers of early pluripotent NCC lineages such as the chromatophore precursor. This hypothesis remains to be tested.

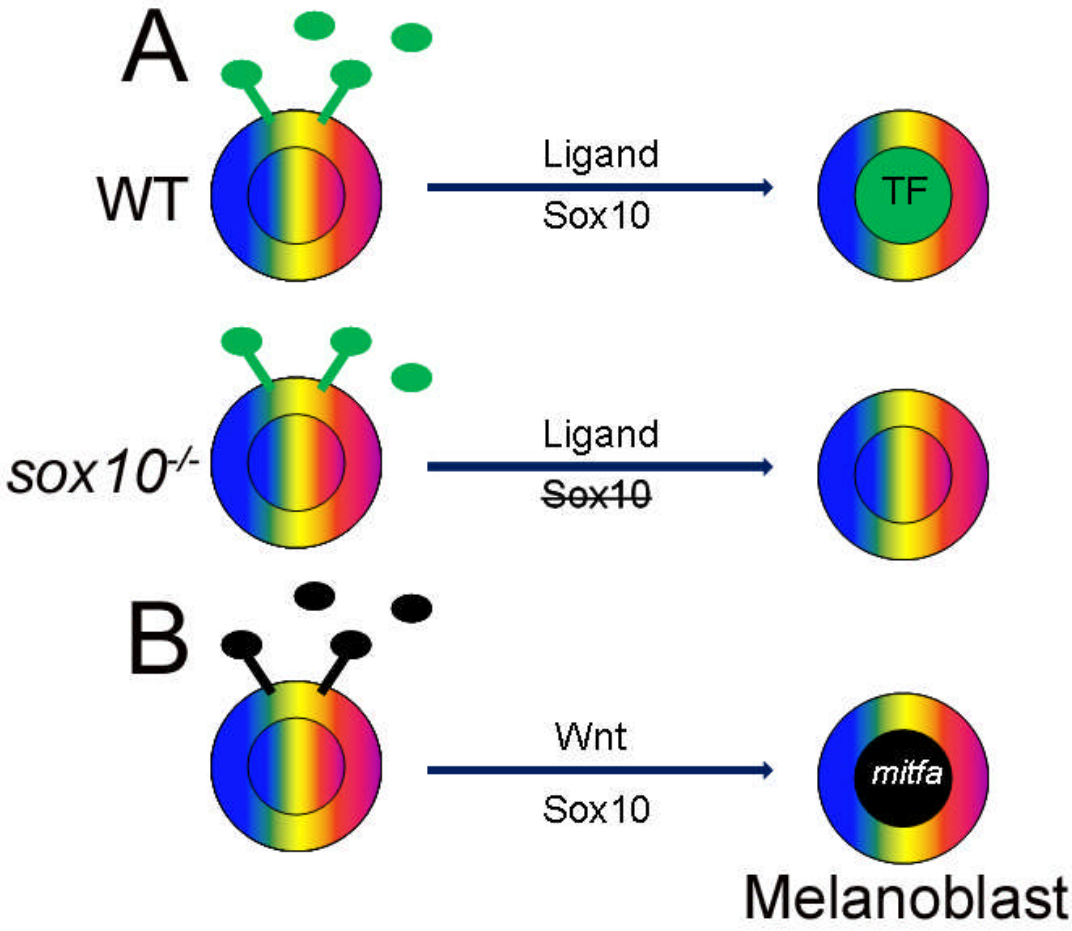


Figure 4: General model of *sox10* and NCC specification.
(A) A schematic of the general role of Sox10 in NCC development. A multipotent NCC (multicoloured shading) becomes specified by an instructive ligand (oval) binding a receptor (stick in cell membrane) in combination with Sox10 signalling. This activates a key downstream TF and the cell becomes specified (indicated by single colour of nucleus). In a *sox10* mutant NCCs a lack of Sox10 signalling results in a failure of specification leading eventually to cell death. (B) Schematic of the general model applied to melanocyte specification. Wnt signalling in combination with Sox10 activates the master switch TF Mitfa to specify melanoblasts. Figure adapted from Kelsh 2006.

1.3.4.4 Sox10 and Glia

In contrast to other NCCs in which *Sox10* expression is transient, *Sox10* expression persists in differentiated glia even into adulthood (Bondurand et al., 1998). This suggests that *Sox10* may play a role in both glial fate specification and differentiation. Homozygous mouse *Sox10* mutants and zebrafish *sox10* mutants display strong defects in glial development thus indicating that *Sox10* is absolutely required by this cell lineage (Britsch et al., 2001, Carney et al., 2006). In *Sox10* mutants, expression of the Neuregulin receptor *ErbB3* in glial progenitors is lost (Britsch et al., 2001). This signalling system plays a key role in glial specification thus loss of *ErbB3* expression in *Sox10* mutants may cause the failure of glial specification (Britsch et al., 2001). *Sox10* mutant NCCs grown in culture conditions that promote gliogenesis cannot adopt a glial fate and expressed markers for

other fates suggesting that fate specification failed (Paratore et al., 2001). While mouse and zebrafish *ErbB3* mutants do show defects in glia, the *Sox10* mutant glial phenotype is more severe (Britsch et al., 2001, Carney et al., 2006). Thus additional mechanisms are likely to contribute to the glial *Sox10* mutant phenotype. Notch signalling is known to play an important role in the fate choice between neurogenesis and gliogenesis with transient activation of notch receptors capable of irreversibly driving cells to a glial fate (Morrison et al., 2000). Interestingly *Notch1* expression is transiently up-regulated in NCCs during early DRG development and this up-regulation fails in *Sox10* mutant embryos (Britsch et al., 2001). Thus at least two mechanisms that promote gliogenesis fail in *Sox10* mutant embryos. Both of these mechanisms promote glial specification through the detection of extrinsic factors by receptors. Thus current evidence indicates that *Sox10* promotes glial specification by regulation of receptors that when lost prevent NCCs adopting a glial fate. However, no master switch TF that drives glial specification has been identified. Therefore the influence of *Sox10* on this gene as per the general model of *Sox10* in specification cannot be evaluated.

As glial specification fails in *Sox10* mutant embryos, the role of *Sox10* in glial differentiation is difficult to assess. However, *Sox10* is known to directly regulate several glial differentiation genes (**Table 1**). The Schwann cell specific protein, protein zero (P_0) is involved in myelin compaction while *Cx32* is expressed in Schwann cells and allows the passage of small molecules through the myelin sheath. Both of these genes are highly expressed in Schwann cells, are required for Schwann cell differentiation and are direct regulatory targets of *Sox10* (Bondurand et al., 2001, Peirano et al., 2000). Indeed, SOX10 up-regulates the expression of both of these genes. Thus *Sox10* has a known function in both glial specification and differentiation. With data from melanocytes suggesting that *Sox10* can regulate *Dct* expression, it will be interesting to identify direct targets of Sox10 to examine the role Sox10 has in differentiating NCCs.

1.3.4.5 Sox10 and otic vesicle development

Human diseases that result from *SOX10* mutations such as WS4 are often associated with deafness. While *Sox10* heterozygous mouse mutants do display behaviour characteristic of hearing problems, the cause and molecular basis of this has yet to be properly characterised. It has been assumed that a loss of melanocytes in the ear, which are NC derived and therefore *Sox10* dependent is the defect that results in hearing impairment. Melanocytes in the stria vascularis of the inner ear are required for both the development of the stria vascularis and for normal function of the inner ear (Tachibana, 1999). At least one function of melanocytes in the stria vascularis involves the secretion of K^+ via ion pumps to help generate the positive potential (endocochlear potential) of the

otic vesicle endolymph. Zebrafish *sox10* mutant embryos display a small otic vesicle and reduced otoliths (Whitfield et al., 1996). Recent work has identified that Sox10 influences ear development independent of the loss of melanocytes in *sox10* mutant embryos (Dutton et al., 2009). Analysis of a number of important otic vesicle patterning genes in WT and *sox10* mutant embryos identified altered expression patterns. In *sox10* mutant embryos expression of patterning genes, for example *fgf8a, fsta* and *bmp4*, all displayed slightly expanded expression domains or ectopic expression (Dutton et al., 2009). Thus the loss of Sox10 function in the ear results in a loss of gene regulation. This suggests at a model for Sox10 function in the ear that is distinct from NC fate specification through the induction of key TFs.

1.3.5 Target genes of Sox10

The function of *Sox10* is becoming understood in a number of NC derivatives. Sox10 is likely to perform these roles by regulating the expression of a number of target genes including master switch TFs. A small number of Sox10 direct targets have been established (**Table 1**). *Sox10* binding sites have been identified in all the promoter regions of these target genes except the β4 subunit of the nicotinic acetylcholine receptor. In addition, as is typical for *Sox* genes, it has been shown that Sox10 acts in synergy with other TFs to achieve optimal rates of target gene transcription in the majority of these cases. For example, co-transfection of *Sox10* and *EGR2* expression constructs with the *Cx32* promoter driving a luciferase reporter gene led to increased activation of the promoter than with either TF independently (Bondurand et al., 2001). The only known direct target of SOX10 that acts as a master switch TF is *Mitf* in the melanocyte lineage (Verastegui et al., 2000). A number of genes have also been observed to be differentially regulated in *Sox10* mutant embryos in comparison to WT embryos (**Table 2**). These genes have not been described as direct targets of Sox10. However in many cases the relationship between Sox10 and these target genes has not been investigated. Of particular interest is whether the TFs in **Table 2** such as *Phox2b, Mash1* or *ngn1* are direct targets of Sox10. These genes have been shown to play a role in the specification of certain NC derivatives and thus are candidate master switch TFs (Carney et al., 2006, Elworthy et al., 2005, Kim et al., 2003). A number of the genes known to be differentially regulated in *Sox10* mutant embryos are also likely to be indirect targets of Sox10. That is, the expression of these genes is perturbed in *Sox10* mutant embryos not as a result of a lack of Sox10 function directly but via incorrect regulation of a target of Sox10. These indirect targets may include genes responsible for cellular differentiation and survival and for NCC processes such as migration.

Tissue/Cell Type	Gene	Evidence	Reference
Glia	*P_0*	Transient transfection reporter assay and Sox10 binding sites	(Peirano et al., 2000)
	Cx32	Transient co-transfection reporter assay and Sox10 binding sites	(Bondurand et al., 2001)
	Mbp	Transient co-transfection reporter assay and Sox10 binding sites	(Wei et al., 2004)
	Cntf	Transient co-transfection reporter assay and Sox10 binding sites	(Ito et al., 2006)
	Plp	ChIP and Sox10 binding sites	(Lee et al., 2008)
	Ptn	ChIP and Sox10 binding sites	
	Sod3	ChIP, Sox10 binding sites transient co-transfection reporter assay and EMSA	
	Sox10	ChIP and Sox10 binding sites	
Neurons of PNS	*c-Ret*	Transient co-transfection reporter assay and Sox10 binding sites	(Lang et al., 2000)
	Ednrb	Sox10 binding sites and EMSA	(Zhu et al., 2004)
	α3 subunit of nACh	Transient co-transfection reporter assay and Sox10 binding sites	(Liu et al., 1999)
	β4 *subunit* of nACh	Transient co-transfection reporter assay and EMSA	
Melanocytes	*Mitf*	Transient co-transfection reporter assay and Sox10 binding sites	(Bondurand et al., 2000, Elworthy et al., 2003)
	Dct	Transient co-transfection reporter assay, Sox10 binding sites and EMSA	(Ludwig et al., 2004, Jiao et al., 2004)
	Tyrp1	Transient co-transfection reporter assay and Sox10 binding sites	(Murisier et al., 2006)
	Tyr	Transient co-transfection reporter assay and Sox10 binding sites	(Murisier et al., 2007)

Table 1: Direct Targets of Sox10.
Abbreviations used: nACh = nicotinic acetylcholine receptor, ChIP = Chromatin immunoprecipitation assay, EMSA = Electrophoretic mobility shift assay.

Cell/Tissue type	Gene	Change of expression level in S*ox10* mutant embryos	Reference
PNS	*ngn1*	Down	(Carney et al., 2006)
	Phox2b	Down	(Elworthy et al., 2005, Kim et al., 2003}
	Phox2a	Up	(Kim et al., 2003)
	Mash1	Down	
Glia	*ErbB3*	Down	(Britsch et al., 2001)
	Hes5	Down	
	Notch1	Down	
	foxd3	Down	(Kelsh et al., 2000)
Melanocyte	*kita*	Down	(Dutton et al., 2001a)
	c-Kit	Down	(Britsch et al., 2001)
Xanthophore	*pax7*	Down	(Hammond et al., 2007)
	gch	Down	(Pelletier et al., 2001)
Iridophore /chromatoblast	*ltk*	Up	(Lopes et al., 2008)
Neural Crest	*crestin*	Down	(Elworthy et al., 2005)

Table 2: Table of Genes differentially regulated in *Sox10* mutant embryos.
The genes present in this table have not been tested and described as direct targets of Sox10.

1.4 Project Overview and Aims

While a number of Sox10 target genes have been identified, this complement of direct and indirect targets is not sufficient to ensure the correct development of all non-skeletogenic NC derivatives. Targets that remain to be identified include master switch TFs for cell lineages such as iridophores and glia, receptors for extracellular instructive cues that drive processes such as fate choice and genes involved in cell differentiation, migration and survival. The identification of such genes will provide key information to describe the molecular process that NCCs are required to undergo during development. As NCCs provide models of both cell development and cancer biology the understanding of these processes is crucial. In addition, a study to identify key genes in NC development will shed light on the underlying causes of a group of human diseases known as neurocristopathies.

The ultimate aim of this study is to identify direct and indirect targets of Sox10. To address the paucity of Sox10 target genes, a microarray approach was adopted. This approach was designed to compare the expression profiles of *sox10* expressing cells from both WT and *sox10* mutant embryos. Towards this end the initial aims were:

1. To isolate GFP positive cells from both WT and *sox10* mutant embryos at 24 hpf from *sox10*:GFP transgenic embryos by fluorescence activated cell sorting (FACS). Subsequently RNA needed to be extracted from these cells suitable for microarray analysis.
2. Analysis of the microarray data is required to identify genes differentially regulated in *sox10* mutant embryos, these genes are candidate direct or indirect targets of Sox10.

3. To validate candidate genes by *in situ* hybridization. This approach will reveal both temporal and spatial information regarding expression of candidate genes in both WT and *sox10* mutant embryos.

Genes whose expression is positively regulated by Sox10 will appear to be down-regulated in *sox10* mutant embryos. These genes should include a number of well known Sox10 target genes that are expressed at 24 hpf such as *mitfa* and *dct*. These down-regulated genes are required to drive the development of NCCs. While a number of genes are known to be down-regulated in *sox10* mutant embryos, very few genes are known to be up-regulated. It is predicted that genes up-regulated in *sox10* mutant embryos will be markers of early NCCs and may include markers of NCSCs.

The targets identified will be limited to the cell types that are defective in *sox10* mutant embryos at 24 hpf such as glia and pigment cells. Specification of other non-skeletogenic NC derivatives such as the ENS and sensory neurons occurs later than 24 hpf therefore differentially regulated marker genes of these lineages are unlikely to be detected.

It was also initially hoped that a gene identified by this process as a particularly interesting target of Sox10 would be further examined through functional studies. In particular it was envisioned that a morpholino study would be utilised to examine the effects of knocking out this specific gene of interest. Due to time constraints this was not completed.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Chemicals

General laboratory chemicals were of analytical research grade from a range of manufacturers. Most chemicals came from Sigma Chemical Company (St. Louis, MO, USA) or Fischer Scientific UK Ltd. (Loughborough, Leicester, UK). Specialist reagents and sources are listed below:

Agarose	Invitrogen (Carlsbad, CA, USA)
Ethidium Bromide	Sigma Chemical Company (St. Louis, MO, USA)
Luria Broth	Sigma Chemical Company (St. Louis, MO, USA)
Luria Agar	Sigma Chemical Company (St. Louis, MO, USA)
Carbenicillin	Bioline (London, UK)
Fetal Bovine Serum	Sigma Chemical Company (St. Louis, MO, USA)
Horse Serum	Sigma Chemical Company (St. Louis, MO, USA)
Kanamycin	Sigma Chemical Company (St. Louis, MO, USA)
6 x Loading Dye	Promega (Madison, WI, USA)
Phenol Red	Sigma Chemical Company (St. Louis, MO, USA)
Sheep Serum	Sigma Chemical Company (St. Louis, MO, USA)
IPTG	Melford Labs Ltd. (Suffolk, UK)
Phenol/chloroform/isoamyl alcohol	Sigma Chemical Company (St. Louis, MO, USA)
1-phenyl-2-tiourea (PTU)	Sigma Chemical Company (St. Louis, MO, USA)
Tricaine (methylsulfonate)	Sigma Chemical Company (St. Louis, MO, USA)
TriReagent™	Sigma Chemical Company (St. Louis, MO, USA)
X-gal	Sigma Chemical Company (St. Louis, MO, USA)
DEPC	Sigma Chemical Company (St. Louis, MO, USA)

2.1.2 Nucleic Acids

DIG RNA labelling mix	Roche Diagnostics GmbH (Mannheim, Germany)
100 bp DNA ladder	Promega (Madison, WI, USA)
1 kb DNA ladder	Promega (Madison, WI, USA)
Primers (see **Table 3**/appendix)	Invitrogen (Carlsbad, CA, USA) or Bioneer (Alameda, CA, USA)

2.1.3 Enzymes

Restriction endonucleases and appropriate 10 x restriction buffers were purchased from Promega (Madison, WI, USA) or New England Biolabs (Beverly, MA, USA). Other enzymes were obtained from the following sources along with appropriate buffers.

Proteinase K	Roche Diagnostics GmbH (Mannheim, Germany)
Trypsin	Invitrogen (Carlsbad, CA, USA)
Pronase (Protease Type XIV)	Sigma Chemical Company (St. Louis, MO, USA)
RNase Inhibitor	Roche Diagnostics GmbH (Mannheim, Germany)
T4 DNA ligase	Promega (Madison, WI, USA)
Go*Taq* DNA polymerase	Promega (Madison, WI, USA)
iScript Reverse Transcriptase	BioRad (Hercules, CA, USA)
SP6 RNA Polymerase	Roche Diagnostics GmbH (Mannheim, Germany)
T7 RNA polymerase	Roche Diagnostics GmbH (Mannheim, Germany)

2.2 Solutions, Buffers and Media

2.2.1 Solutions and Buffers

Hybridization mix
- 50 % Formamide
- 5 x SSC
- 0.25 mg Heparine
- 2.5 mg TRNA
- 0.1 % Tween20
- 9 mM Citric Acid
- In Sterile Water

Holtfreter's solution
- 0.35 % (w/v) NaCl
- 0.005 % (w/v) KCl
- 0.01 % (w/v) $CaCl_2$
- 0.02 % (w/v) NaHC03
- In MilliQ Water

In situ blocking buffer
- 5 % Sheep serum
- 2 mg/ml BSA
- 0.1 % Tween
- In PBS

NBT/BCIP buffer
- 100 mM Tris-HCl pH 9.5
- 50 mM $MgCl_2$
- 100 mM NaCl
- 0.1 % Tween20
- In MilliQ Water

PFA
- PBS with 4 % Paraformaldehyde

PBS
- 2.7 mM KCl
- 137 mM NaCl
- 0.01 M Phosphate buffer (Na_2HPO_4 + KH_2PO_4) pH 7.4

PBT
- PBS with 0.1 % Tween20

1 x SSC
- 0.15 M Sodium chloride
- 0.015 M Sodium citrate
- Adjust to pH 7.2 with NaOH

1 x TAE
- 40 mM Tris-HCl
- 20 mM Sodium acetate
- 2 mM EDTA
- Adjust to pH 7.8 with glacial acetic acid

5 % Fetal Bovine Serum (FBS)
- 5 % FBS in PBS

2.2.2 Media

Embryo Medium

0.50 µM NaCl
0.17 µM KCl
0.33 µM CaCl
0.33 µM $MgSO_4$
0.1 % methylene blue

LB

2.5 % (w/v) Luria broth base

LB-agar

3.7 % (w/v) Luria agar base

2.3 All other Reagents

2.3.1 Plasmids

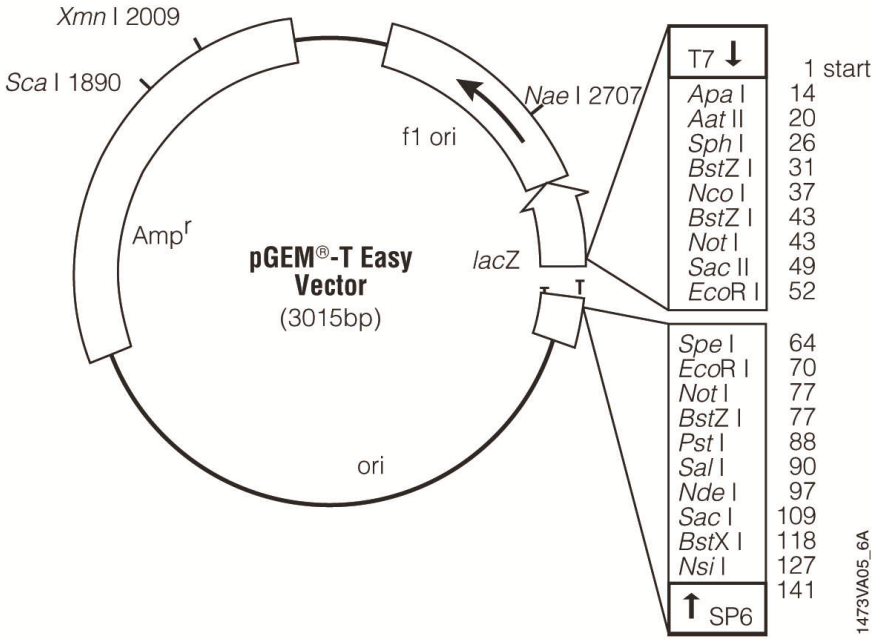


Figure 5: Vector map of pGEM-T® easy.
Map shows the ampicillin resistance gene, the restriction enzyme cutting sites in the multiple cloning region and the transcription initiation sites for SP6 and T7 flanking the multiple cloning region. Taken from the Promega pGEM-T® easy manual. Additionally plasmids for generating some *in situ* probes were very kindly sent by the following people:

Otomp – Tanya Whitfield
Pax7 – Hitoshi Okamoto
Wnt11r - Simon Hughes
Hoxd4a – Victoria Prince

2.3.2 Primers

Custom primers were purchased from Invitrogen (Carlsbad, CA, USA) or Bioneer (Alameda, CA, USA). Primers in Table 3 were used to amplify key target genes from cDNA, for a full list of primers used to amplify targets identified by microarray analysis see appendix.

Primer Name	Sequence	Expected Band Size (bp)
ActinF	GGTATGGGACAGAAAGACAG	330 (680 from genomic DNA)
ActinR	AGAGTCCATCACGATACCAG	
DctF	AACTGTGGCGAGTGTAAGTT	697
DctR	GACTGTCGAAGTCTCTGAGG	
MitfaF	GCTTTCCAGTAGAAGCAGAA	581
MitfaR	GTTGTCCATAAGCATGTCCT	
Sox10F (S21)	AACTACCGAAGTCACCTGTGG	420
Sox10R (S22)	GATATTGATCCGCCAGTTTCC	
OtompF	GTTTTCACAGTCTTTGTATG	355
OtompR	AGTCTACGACAAATAACACA	
GFPf	ATGGCCAACACTTGTCAC	600
GFPr	TGCTAGTTGAACGCTTCC	

Table 3: Primers.
All primer pairs were designed to amplify a fragment of the named gene from zebrafish cDNA. Primer sequences are given in the 5' to 3' direction of their respective DNA strands.

2.3.3 Kits

iScript™ cDNA synthesis kit	Bio-Rad, Hercules, CA, USA
Plasmid Midi kit	Qiagen, Valencia, CA, USA
RNeasy micro kit	Qiagen, Valencia, CA, USA
Wizard® *Plus SV* miniprep DNA purification	Promega, Madison, WI, USA
QIAquick® Gel extraction kit	Qiagen, Valencia, CA, USA
DIG RNA labelling kit	Roche Diagnostics GmbH, Mannheim, Germany
Ovation Biotin RNA Amplification and Labelling system.	NuGEN, San Carlos, CA, USA

2.3.4 Microarray Chips

Affymetrix GeneChip® Zebrafish Genome Array – Affymetrix, Inc, Santa Clara, CA, USA.

2.3.5 Software

Photoshop™ 7 (Adobe) was used to analyse and enhance captured images and to compile and annotate figures. Photoshop™ CS2 (Adobe) was used to merge images to create composite pictures of whole embryos. Microsoft Excel 2003 or 2007 was used to record results from analyses of the microarray data and to perform basic statistical tests. Microsoft Word 2007 was used for word processing. Initial Microarray assessment and initial analysis were performed using the Affymetrix GeneChip Operating Software

(GCOS) available freely on www.Affymetrix.com. Array data normalization and identification of differential gene expression was performed using the Gene Expression Pattern Analysis Suite (GEPAS) freely available on the WWW at www.gepas.org.

2.4. General Techniques

2.4.1 Transformation of Competent *E. coli*

An appropriate amount of plasmid DNA was introduced into 50 µl of thawed chemical competent XL-1 Blue competent cells (Stratagene, Cedar Creek, TX, USA). Transformation was performed using a standard heat shock protocol following manufacturer's guidelines. Post transformation, cells were incubated in LB shaking at 37 ºC for one hour followed by spreading on LB-agar plates containing the appropriate antibiotic and incubated at 37 ºC overnight. When required, blue white selection was performed by addition of IPTG and X-gal to concentrations of 0.016 mg/ml and 0.04 mg/ml respectively.

2.4.2 Purification of Plasmid DNA

Plasmid DNA was prepared from bacterial suspensions, inoculated using freshly transformed cells, supplemented with the appropriate antibiotic using Promega miniprep Wizard™ kits as per the manufacturer guidelines.

2.4.3 Restriction Digests

Digestion of DNA was performed using conditions specified by the enzyme manufacturer for at least one hour at the appropriate incubation temperature. Five units of each restriction endonuclease were added per 1 µg of DNA while ensuring the volume of enzyme never exceeded 10 % of the total digest volume.

2.4.4 DNA Sequencing

DNA sequencing was performed by MWG-Biotech (Ebersberg, Germany). 2-3 µg of DNA was precipitated per sequencing reaction using 0.1 volumes of 8 M LiCl and 2.5 volumes of ice-cold 100 % ethanol and incubated on ice for 30 minutes. Samples were centrifuged at 14,000 rpm in a Techne Genofuge 16M desktop centrifuge, washed in 70 % ethanol and air-dried. Alternatively 15 µl of liquid plasmid DNA (concentration 100-150 ng/ µl) was sent for sequencing as specified on the MWG-Biotech website (www.mwg-biotech.com). Sequencing was performed from the T7 RNA transcription start site of the pGEM-T easy vector (**Figure 5**). Sequence results were analysed using the NCBI BLAST tool on the World Wide Web (WWW) to confirm cloning of part of the correct gene and to identify insert orientation. The gene fragment cloned was further analysed for restriction enzyme sites using NEB cutter v2.0 on the WWW (http://tools.neb.com/NEBcutter2/index.php).

2.4.5 Electrophoresis of DNA and RNA

Size fractionation of DNA was performed by electrophoresis on agarose (1-2 %) gels supplemented with ethidium bromide, in TAE running buffer. Samples were loaded with 1x gel loading buffer, electrophoresed at 90 V and visualised using UV light. Images were captured using an Alphaimager™ 3400. RNA electrophoresis was performed as stipulated above under RNase free conditions using RNase free agarose and TAE running buffer made using DEPC water. The gel tray, comb and gel tank were soaked in 1 % Hydrogen peroxide prior to pouring and running the gel.

2.4.6 Phenol Chloroform extraction of DNA

DNA solutions were made up to 100 µl with 1 x TE and 1 volume of Phenol:Chloroform:Isoamyl alcohol (Sigma-Aldrich, St. Louis, MO, USA) was then added. After vigorous mixing the sample was centrifuged for 5 minutes at 14,000 rpm on a desktop centrifuge (Techne Genofuge 16M). The upper, aqueous, layer was collected and contaminating phenol was removed by washing with an equal volume of chloroform followed by centrifugation as above. DNA was recovered by ethanol precipitation

2.4.7 Ethanol Precipitation of DNA

DNA samples were precipitated by addition of sodium acetate pH 5.2 to a final concentration of 0.3 M and 2 volumes of ice cold ethanol. To allow the DNA to precipitate samples were left on ice for 30 minutes, DNA was recovered by centrifugation, 14,000 rpm for 15 minutes, on a desktop centrifuge (Techne Genofuge 16M). The pellet was washed with ice cold 70 % ethanol, air dried at room temperature and resuspended using nuclease free water.

2.4.8 Polymerase Chain Reaction (PCR)

PCRs were performed in 25 µl reactions using GoTaq Green PCR master mix (Promega, Madison, WI, USA). Appropriate primer concentrations, 0.1-1 µM, and template DNA, <250 ng, were used. Amplification was performed in either a Techne Touchgene (Cambridge, UK) or a G-Storm (Gene Technologies Ltd, Braintree, Essex, UK) thermal cycler. Programs consisted of an initial template denaturation at 95 ºC for 5 minutes, 35-40 cycles consisting of denaturation at 94 ºC for 30 seconds, 30 second primer annealing step at a temperature appropriate for the primers and an extension step at 72 ºC for 30 seconds. A final extension step of 5 minutes at 72 ºC was included. To determine optimal primer annealing temperatures the primer annealing step was sometimes replaced with a gradient PCR step thus allowing multiple annealing temperatures to be tested.

2.4.9 Reverse Transcription

Reverse transcription was performed using the iScript™ cDNA synthesis kit (Bio-Rad, Hercules, CA, USA) following manufacturer's guidelines. Briefly, 1 µg of RNA was added along with iScript Reverse Transcriptase and reaction mix and then the following heat

steps were performed in either a Techne Touchgene (Cambridge, UK) or a G-Storm (Gene Technologies Ltd, Braintree, Essex, UK) thermal cycler, 5 minutes at 25 ºC, 30 minutes at 42 ºC, 5 minutes at 85 ºC. Quality of cDNA was assessed using Actin, Dct and Sox10 primers (**Table 3**).

2.4.10 PCR product purification

A PCR reaction was size fractionated by gel electrophoresis on a 1 % agarose gel. The gel was visualised under UV lighting wearing appropriate protection on a UV transilluminator (UVP, Upland, CA, USA). A glass plate was placed between the gel and the UV source and the duration of UV exposure was kept to a minimum. Overexposure of DNA to UV light can result in the formation of pyrimidine dimers which can lead to very few or no positive ligations occurring. A band at the correct size for the primer pair used during PCR was excised using a scalpel blade and DNA was purified from the agarose using the QIAquick® Gel extraction kit (Qiagen, Valencia, CA, USA) following manufacturer's guidelines for the microcentrifuge protocol. A small amount of eluted DNA was visualized using gel electrophoresis to confirm gel extraction had worked.

2.4.11 pGEMT®-easy Ligations

Ligation of purified PCR products into pGEMT®-easy vectors (Promega, Madison, WI, USA) was performed as per the manufacturer's guidelines. Briefly 1 µl of purified PCR product was mixed with vector and T4 DNA ligase (Promega, Madison, WI, USA) and incubated over night at 4 ºC. Appropriate positive and background controls were performed to assess the efficacy of the ligation reaction. Ligations were transformed into competent bacteria.

2.4.12 DNA and RNA quantification by spectrophotometry

To quantify the concentration of DNA and RNA samples a Beckman DU 530 spectrophotometer (Beckman Coulter Inc., Fullerton, CA, USA) was used. Samples, appropriately diluted, were assayed using light absorbance at a wavelength of 260 nm (A_{260}). Typically a sample of DNA purified by miniprep was diluted 1:20 using nuclease free water, RNA was diluted with an appropriate amount of DEPC treated water. Additionally RNA sample absorbance was assayed at 280 nm (A_{280}) wavelength to compare the ratio between A_{260} and A_{280}, this provided a measure of purity.

2.4.13 Total RNA extraction from whole embryos

RNA work must be carried out under RNase free conditions. 150 dechorionated embryos at the appropriate developmental stage suspended in 1 ml of TriReagent™ (Sigma Chemical Company, St. Louis, MO, USA) were homogenized by repeatedly passing through a needle (21Gx37) attached to a syringe. The homogenate was centrifuged for 10 minutes at 14,000 rpm in a Techne Genofuge 16M desktop centrifuge. The supernatant was decanted and 0.2 ml of chloroform was added followed by vigorous mixing and

incubated for 15 minutes. After incubation the sample was centrifuged for 15 minutes at 14,000 rpm in a Techne Genofuge 16M and the RNA containing upper aqueous phase was transferred to a new tube. RNA was precipitated by addition of 0.5 ml isopropanol, incubation for 5-10 minutes and by centrifugation for 10 minutes at 12,000 rpm on a microcentrifuge. The supernatant was then removed and the RNA pellet was washed in 75 % ethanol. Ethanol was removed and the pellet allowed to air dry quickly. The pellet was resuspended in 10 µl of DEPC-treated water. 1 µl of RNA was visualized by gel electrophoresis and 1µl was quantified by spectrophotometry. Remaining RNA was stored at -20 ºC short term or -80 ºC long term.

2.5 Zebrafish Techniques

2.5.1 Fish Husbandry

Embryos were obtained through appropriate crosses of wild type, mutant and transgenic fish stocks maintained in the University of Bath zebrafish facility. Embryos were raised in embryo medium and staged according to Kimmel et al., 1995. Dechorionation was performed using Dumontstar number 5 biological forceps or by incubation in 2 mg/ml pronase (Protease Type XIV, Sigma Chemical Company St. Louis, MO, USA) dissolved in embryo medium for appropriate time periods, typically 10-15 minutes. Embryos were anaesthetised with tricane (Sigma Chemical Company St. Louis, MO, USA) diluted to approximately 0.2 % final volume prior to manipulation, live mounting or fixation when appropriate. Fixation was performed in 4 % PFA overnight at 4 °C. Embryos older than 18 somites were dechorionated prior to fixation. To inhibit melanisation embryo medium was supplemented with 0.003 % 1-phenyl-2-thiourea (PTU) when required. Sections were taken of resin embedded embryos (see Appendix for protocol).

2.5.2 Injection of zebrafish embryos

Injection needles were made by pulling Bores Oocyte Inj 90 mm flared glass capillaries (Drummond Scientific Company, Broomall, PA, USA) on a Micropipette puller (Sutter Instrument Company, Novato, CA, USA). Embryos were pipetted onto a 2 % agarose plate and held in a groove in the agarose. Injections were performed under a dissecting microscope using a Drummond Nanoinject II apparatus (Drummond Scientific Company, Broomall, PA, USA). Phenol Red was added to the injection solution and dispensed from needles backfilled with mineral oil.

2.5.3 Embryo mounting and microscope techniques

Low power microscopic analysis and general embryo manipulation was performed on dissecting microscopes, either a Leica MZ7.5 or a Leica MZFLIII for fluorescence work. For detailed analysis fixed embryos were mounted between No. 1 cover slips in 50 %

glycerol and examined on an Eclipse E800 (Nikon) microscope using Nomarski optics. Images of embryos were captured on a Digital Sight DS-5M camera with U1 Controller (Nikon). Cell counts were performed using a haemocytometer and examined on an Eclipse E800, dead cells were stained fluorescently using 7-Amino-actinomycin D (7-AAD).

2.5.4 *In situ* Hybridization of whole mount embryos

In situ probes were synthesized under sterile RNase free conditions. 10 µg of DNA was linearized by digesting 5' of the probe sequence with an appropriate restriction enzyme. The reaction was stopped and DNA purified by phenol:chloroform extraction followed by ethanol precipitation. DIG labelled probes were synthesized using a DIG RNA labelling kit (Roche GmbH, Mannheim, Germany) following manufacturer's guidelines. Briefly, approximately 1 µg linearized DNA was added to 1 µl DIG labelled NTP mix, 2µl transcription buffer, 1µl RNase Inhibitor and made up to a total volume of 18 µl with DEPC treated MilliQ water. Finally 2µl of the appropriate RNA polymerase was added and the reaction was incubated at 37 ºC for 2 hours. RNA was purified by adding 2.5 µl 4 M LiCl and 70 µl 70 % cold ethanol and incubated overnight at -20 ºC. RNA was pelleted by centrifugation at 14,000 rpm for 15 minutes on a desktop centrifuge (Techne Genofuge 16M). The pellet was washed with 70 % cold ethanol then briefly air dried and dissolved in 10 µl RNase free water. The probe was visualised by electrophoresis of 1 µl on an agarose gel. The rest of the probe was made up to 100 µl using hybridization mix.

Fixed embryos were dehydrated by washing three times in 100 % methanol (MeOH) followed by incubation at -20 ºC for a minimum of 30 minutes. Embryos were rehydrated by washing in 75 % MeOH 25 % PBS, then 50 % MeOH 50 % PBS, then 25 % MeOH 75 % PBS followed by four PBT washes for 5 minutes each, all PBT washes were 5 minutes unless otherwise stated. Embryos older than 18 somites were digested with Proteinase K, 1:1,000 or 1,10,000 dilution, for an appropriate period of time. Embryos were washed quickly in PBT and refixed for 20 minutes in 4 % PFA followed by 4 PBT washes.

The following steps were performed typically at 65 ºC or occasionally at 68 ºC to improve hybridization specificity if appropriate for the probe. Embryos were washed in hybridization mix (HM) at 65 ºC for 15 minutes and then incubated in HM for 2-5 hours. Embryos were incubated overnight in 200 µl of diluted DIG labelled probe, typically 1:100 or 1:200 in HM, to hybridize the probe. Subsequently embryos were washed quickly in HM followed by 10 minute washes in 75 % HM 25 % 2 x SSC, 50 % HM 50 % 2 x SSC, 25 % HM 75 % 2 x SSC and 2 x SSC. Embryos were then washed in 0.2 x SSC twice for 30 minutes at 65 ºC. All further washes were performed at room temperature. Embryos were washed for 5 minutes in 75 % 0.2 x SSC 25 % PBT, 50 % 0.2 x SSC 50 % PBT, 25 % 0.2

x SSC 75 % PBT then PBT. Hybridized embryos were blocked by incubation *in situ* blocking buffer for 2-4 hours. Blocked embryos were incubated with anti-DIG alkaline phosphatase conjugated Fab fragment (Roche GmbH, Mannheim, Germany) diluted 1:5,000 with *in situ* blocking buffer at 4 ºC overnight with gentle agitation.

The antibody was removed and embryos were washed once quickly in PBT followed by 6 PBT washes then three 5 minute washes in NBT/BCIP buffer. Embryos were transferred to glass nine well plates and incubated in NBT/BCIP blue staining stock solution (Roche GmbH, Mannheim, Germany) diluted 1:50 in NBT/BCIP buffer. Staining reaction was stopped by a fast PBT wash followed by a 5 minute PBT wash. Alternatively to reduce non-specific staining embryos were washed quickly in PBT followed by a wash in 100 % EtOH for an appropriate time period then washed in PBT. Embryos were stored at 4 ºC in 50 % glycerol.

2.5.5 FACS sorting of zebrafish embryos

To generate *sox10* mutant embryos expressing GFP under the control of the *sox10* promoter (*Tg(-7.2sox10*:EGFP);*sox10*$^{m618/m618}$ embryos), heterozygous *sox10*$^{+/m618}$ fish were crossed with *Tg*(*-7.2sox10:*EGFP);*sox10*$^{+/m618}$ fish. This cross generated *sox10* mutant and wild type embryos in a normal Mendelian ratio (1:3) and 50 % of the embryos expressed GFP. At 20 hpf embryos were sorted for GFP expression under a dissecting microscope. Approximately 500 GFP positive embryos were then dechorionated. Removal of the chorion reduced the number of GFP negative events that were counted during cell sorting. Embryos were then lightly anaesthetized and sorted into *sox10* mutant and WT GFP positive groups. 500 GFP positive embryos were selected to ensure that at least 100 mutant embryos could be sorted based on normal Mendelian genetics. *Tg*(*-7.2sox10*:EGFP);*sox10*$^{m618/m618}$ embryos were identified from non-mutant transgenic embryos by visualizing differences in GFP expression at 22 hpf. At this time point *Tg*(*-7.2sox10*:EGFP);*sox10*$^{m618/m618}$ embryos showed reduced GFP expression in the ear compared to WT embryos. Mutant embryos also displayed a clumping of GFP expression in the head close to the midbrain/hindbrain boundary, in WT embryos GFP expression was more evenly spread. 100 *Tg*(*-7.2sox10*:EGFP);*sox10*$^{m618/m618}$ embryos and 100 WT embryos in separate 1.5 ml eppendorf tubes were washed three times in 1x Holtfreter's solution and then disaggregated at 24 hpf with a micropestle (Eppendorf, Hamburg, Germany) by firmly inserting the micropestle into the tube 5 or 6 times. Disaggregated embryos were pelleted at 3,000 rpm for 5 minutes in a desktop centrifuge (Techne Genofuge 16M), this helped to separate cells from yolk material. The supernatant was removed and the pellet was resuspended in 1 ml 0.05 % trypsin plus 53 mM EDTA (Invitrogen, USA) then incubated at 33 ºC for 35 minutes with gentle mixing by pipetting with a P1000 Gilson pipette every 5 minutes. Mixing helped to break up large clumps of cells and enabled small clumps to be broken down into single cells. Cells were pelleted by

spinning at 4,000 rpm, 4 ºC for 5 minutes in a desktop centrifuge (Jencons-PLS Spectrafuge 24D). The pellet was washed three times in PBS to remove the trypsin and then resuspended in 500 µl of ice cold 5 % FBS in PBS and filtered using 40 µm cell strainer (BD Biosciences, NJ, USA) into a 5 % FBS coated chilled 50 ml falcon tube (BD Bioscience, NJ, USA). The addition of 5 % FBS helped to prevent cells sticking to the plastic equipment. 500 µl of 5 % FBS was used to rinse out the falcon tube and then passed through the filter. The filter was rinsed through with another 500 µl of 5 % FBS. The samples were filtered to ensure that only single cells were taken for FACS.

To enrich for GFP positive cells, cells from both *sox10*$^{m618/m618}$ and WT embryos were sorted independently for GFP and collected in RLT buffer (Qiagen Ltd, West Sussex, UK) containing 0.1 % β-Mercaptoethanol on a BD FACSVantage^{TM} SE cell sorter system (BD Bioscience, NJ, USA) into FACS tubes (BD Bioscience, NJ, USA). Samples were transferred into FACS 5ml polypropylene round bottom tubes (BD Bioscience, NJ, USA) coated with 5 % FBS for loading onto the FACS machine. The FACS machine detector amp was set as follows, FSC-H typically to 8, SSC-H set to log and FL1-H set to 700. The laser used was an argon 488 nm laser capable of exciting GFP. Gates identifying the cell populations to be sorted were set on a 50,000 event run using FACStation^{TM} software. 500,000 events were subsequently collected 4 times, with each run taking between 8-10 minutes. Between each run the collected samples were mixed by pipetting to ensure all sorted cells were mixed with RLT. Collected samples were stored at -80 ºC prior to RNA extraction.

To enrich for NCCs from *Tg(-4.9sox10*:EGFP) fish the protocol above was followed except for the identification of WT from mutant fish using the pattern of GFP expression at 24 hpf. *Tg(-4.9sox10*:EGFP) embryos only displayed weak ear expression so the embryos were genotyped using the pattern of GFP positive cells in the head only, WT fish displayed an even spread of GFP positive cells over much of the head while mutant embryos displayed an obvious clumping of GFP expressing cells.

2.6 Preparing Samples for Microarray analysis

2.6.1 RNA Extraction and cDNA synthesis

RNA was extracted from FACS sorted samples using an RNeasy Micro kit (Qiagen Ltd. West Sussex, UK), one column was used per sample. 12 µl of sample was eluted per column; each sample was made up to 50 µl with DEPC water. 10µl was utilised as a template to generate cDNA with an iScript^{TM} cDNA synthesis kit (Bio-Rad, Hercules, CA, USA) following the manufacturer's guidelines. WT samples versus *sox10*$^{m618/m618}$ samples were checked for differential gene expression by PCR using *actin, sox10, dct*, *mitfa* and

otomp primers. RNA samples were sent on dry ice to the University of California, Irvine (UCI), DNA and Protein Microarray Facility for processing and hybridization to an Affymetrix Genechip Zebrafish Genome Array.

2.6.2 RNA Sample Preparation and hybridization to the array

Eluted total RNAs were quantified with a portion of the recovered total RNA adjusted to a final concentration of 5-10 ng/µl. All starting total RNA samples were quality assessed prior to beginning target preparation/processing steps by running out a small amount of each sample onto a RNA Lab-on-A-Chip (Caliper Technologies Corp. Mountain View, CA) that was evaluated on an Agilent bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA). Single-Stranded, then double-stranded cDNA was synthesized from the poly(A) + mRNA present in the isolated total RNA followed by linear amplification, fragmentation and labelling using the Ovation™ Biotin Pico RNA Amplification and Labelling system and the FL - Ovation™ cDNA Biotin Module (NuGEN, San Carlos, CA) as per manufacturer's guidelines. 3.75 µg of the resulting single-stranded cDNA generated during the amplification process was fragmented and labelled to an average strand length of 85 bases (range 80-100 bases) following prescribed protocols (NuGEN, San Carlos, CA). Subsequently, 3.4µg of this fragmented target cDNA was hybridized at 45°C with rotation for 18 hours (Affymetrix GeneChip® Hybridization Oven 640) to probe sets present on an Affymetrix Zebrafish array. The GeneChip® arrays were washed and then stained (SAPE, streptavidin-phycoerythrin) on an Affymetrix Fluidics Station 450, followed by scanning on a GeneChip® Scanner 3000. The results were quantified and initially analyzed, including checking all appropriate control probes on each array, using GCOS v1.4 software (Affymetrix Inc.) using default values (Scaling, Target Signal Intensity = 500; Normalization, All probe Sets; Parameters, all set at default values).

2.6.3 Analysis of Array data

Analysis of differential gene expression between WT versus *sox10*$^{-/-}$ arrays was performed using the web based free software Gene Expression Pattern Analysis Suite (GEPAS) v3.1 (http://gepas3.bioinfo.cipf.es/). Affymetrix Cell Intensity Files (.cel files) provided by the UCI DNA and Protein Microarray Facility were uploaded in a compressed .zip file format to the Expresso tool on GEPAS for pre-processing. The default settings of Background correction = rma, normalization = quantiles, PM correction = pmonly and summary = medianpolish were used. The output file from Expresso was copied and pasted into a Microsoft notepad file and class information was added to identify the columns of wild type and mutant data. This file was uploaded to the GEPAS T-Rex tool for analysis of differential gene expression using both the T-test and a CLEAR test (significance level = 0.05) tools. The m value output from the CLEAR test equals the signal log ratio; this can be used to calculate fold change as follows:

If Signal Log Ratio ≥ 0 then Fold Change = $2^{\text{Signal Log Ratio}}$

If Signal Log Ratio < 0 then Fold Change = $(-1)*(2^{-(\text{Signal Log Ratio})})$

The output file from the T-Rex analysis was copied and pasted into a Microsoft Excel spreadsheet and additional annotation was added. Gene name and GO term information was taken mostly from the Netaffx™ Analysis Centre (http://www.affymetrix.com/analysis/index.affx). When additional information could be found from the following sources; Zebrafish Information Network (ZFIN) website (http://zfin.org) or the *Danio rerio* Genome build 7 (Zv7) on the Ensembl website (http://www.ensembl.org/index.html), it was added to existing annotation. The Ensembl website included information on both Affymetrix probe identifiers and links to genes on the ZFIN website in a single location and thus provided a useful tool to connect both these resources. If a conflict in annotation was identified, the information provided by Affymetrix was preferred.

2.6.4 Searching microarray data for GO terms of interest.

The web based tool Babelomics v3.1 (http://babelomics.bioinfo.cipf.es) was used to perform functional analysis of the microarray data. This tool was developed by the same group as the GEPAS tool used previously to analyse the microarray data. Affymetrix identifiers of interest were uploaded to the Babelomics FatiGO search tool. The search tool was run using default options except duplicates were never removed. All analyses were performed by examining lists of genes for GO biological process and GO molecular function terms of interest. Affymetrix identifiers associated with a GO term of interest were annotated by uploading the Affymetrix identifiers to a NetAffx batch query.

Chapter 3: Towards an Initial Validation of a Microarray Analysis of *Sox10* targets

3.1 Introduction

In this chapter, the series of experiments that were pursued to generate microarray data for comparing gene expression between wild type (WT) and *sox10*$^{-/-}$ neural crest cells (NCCs) will be presented. It is hoped that the generation of such data will lead to the identification of novel targets of Sox10. Initially Green Fluorescent Protein (GFP) expressing cells were purified from two different zebrafish transgenic lines at approximately 24 hours post fertilization (hpf) by fluorescent activated cell sorting (FACS). Chronologically, the *Tg*(*-7.2sox10*:EGFP) cells were purified first and RNA was extracted from them. The RNA was sent to the Microarray Facility at the University of California, Irvine, for hybridization to Affymetrix zebrafish GeneChips. The resulting data was then analysed for differential gene expression, the quality of this was assessed using some key NC genes and an otic epithelium expressed gene as markers. Cells from the *Tg*(*-4.9sox10*:EGFP) were also purified and subjected to the same process, proceeding a few months later.

3.1.1 Zebrafish Transgenic Lines Utilised in this Project

In this project, two stable zebrafish transgenic lines have been employed that were previously generated in the lab. Both lines express GFP under the control of DNA directly upstream from the ATG start site of the Sox10 coding region. The *Tg*(*-7.2sox10*:EGFP) (Dutton et al., 2008) contains 7.2 kb of the *sox10* promoter region driving GFP while the *Tg*(*-4.9sox10*:EGFP) (Carney, 2003, Carney et al., 2006) line contains 4.9 kb of the *sox10* promoter region driving GFP. The *Tg*(*-7.2sox10*:EGFP) has been published under this nomenclature (Hoffman et al., 2007) and as *Tg*(*-4725sox10*:GFP)^{ba4} (Dutton et al., 2008). For consistency the two lines will be distinguished by the length of the promoter sequence driving GFP and shortened to *7.2sox10*:GFP and *4.9sox10*:GFP. In both lines, GFP expression marking NCCs is first seen around the one somite stage (approximately 11 hpf) and by 24 hpf both lines display strong GFP expression that replicates endogenous *sox10* expression as determined by *in situ* hybridization (Carney, 2003, Carney et al., 2006, Dutton et al., 2008). This includes GFP expression in trunk pre-migratory NCCs, NCCs migrating on the medial and lateral pathways as well as migratory cranial NCCs and glial precursor cells in the head (Carney et al., 2006, Dutton et al., 2008, Carney, 2003). The two lines display a striking difference; *7.2sox10*:GFP embryos display strong GFP expression in the otic epithelium, a site of strong *sox10* expression, while the *4.9sox10*:GFP line displays markedly weaker otic epithelium expression (Carney et al.,

2006, Dutton et al., 2008, Carney, 2003). The expectation when using FACS is that NCCs from both lines will be purified but otic epithelium cells will be strongly reduced in samples from the *4.9sox10*:GFP line compared to the *7.2sox10*:GFP line. Thus microarray data from the *4.9sox10*:GFP line may not identify otic vesicle expressed genes as differentially regulated. By comparing the data from both lines it might be possible to identify otic vesicle expressed genes. Additionally both lines show strong ectopic GFP expression in the most anterior somites (Carney et al., 2006, Carney, 2003) The GFP expression patterns displayed by both these lines is summarised in **Table 4**.

	7.2sox10:GFP	*4.9sox10*:GFP	*sox10 in situ* hybridization
Pre-migratory NC	✓	✓	✓
Migrating NC	✓	✓	✓
Otic Epithelium	✓	✓ (weak)	✓
Cranial NC	✓	✓	✓
Olfactory Neurons	✓	✓	✓
Branchial Arches	✓	✓	×
CNS Brain	×	✓	✓
CNS Neural Tube	✓	✓	✓
Glial Precursors	nd	✓	✓
Schwann Cells	nd	✓	✓
Muscle	✓	✓	×

Table 4: Summary of the sites of GFP expression in the *sox10* reporter transgenic lines at 24 hpf.
Sites where expression has been identified are indicated by a tick, sites where expression has not been noted are identified by a cross. Sites where no information is available regarding expression are identified by nd for not defined. Information collated from (Carney et al., 2006, Dutton et al., 2008, Carney, 2003).

3.1.2 Fluorescence Activated Cell Sorting

Cell sorting by flow cytometry enables the enrichment of a heterogeneous population of cells into defined populations based on detectable cellular properties. A commonly used criterion for cell sorting is cellular fluorescence, either emitted from an expressed fluorescent molecule or from a fluorophore bound to a cell by a specific antibody; this is known as fluorescence activated cell sorting (FACS). For a comprehensive review of flow cytometry and FACS see Practical Flow Cytometry, 4th Ed by H. M. Shapiro which is free to view on the Invitrogen Website. Briefly, a stream of single cells is passed through a

beam of laser light and the character of interest is measured, in the case of this project, GFP expression. The stream is broken into droplets containing a single cell by a vibrating mechanism. If a GFP positive cell is detected voltage is then applied to impart an electric charge to the droplet. The droplet stream subsequently passes through an electric field generated between two deflecting plates with strongly polarized charges. Droplets to be sorted are deflected out of the stream towards the plate with the opposite charge and collected. Uncharged cells are not deflected out of the stream and pass into a waste collector (Shapiro, 2003). The precise nature of the cells to be sorted can be defined with user-selected gates on the software operating the FACS machine. FACS provides a powerful tool to enrich for living cells of a defined population that can then be subjected to further study.

The NC consists of a wide diversity of different cell types, which express a range of cell markers and migrate to many locations within an embryo; they are therefore a difficult cell population to purify by dissection. Flow cytometry has the potential to be a useful tool for sorting NCCs from a disaggregated embryo yet very few examples of this appear in the literature. Indeed when this work commenced there were no examples although FACS has now been employed to enrich for populations of NCCs. Recently cranial neural crest multipotent progenitor cells have been enriched by FACS from a culture of human embryonic stem cells (Zhou and Snead, 2008). Melanocytes expressing GFP under the control of the Fugu *tyrp1* promoter have also been enriched by FACS from transgenic zebrafish embryos (Zou et al., 2006). The zebrafish provides a powerful model organism for the generation of transgenic embryos (Grabher and Wittbrodt, 2008, Kwan et al., 2007, Villefranc et al., 2007) and therefore has the potential to work effectively with FACS to purify cell populations of interest from an *in vivo* source. Indeed there are several recent examples where this has been achieved including enrichment of melanocytes (Zou et al., 2006), *lmo2* positive hematopoietic and endothelial cells (Zhu et al., 2005), *fli1* positive hematopoietic, endothelial and pharyngeal arch cells (Covassin et al., 2006), *vasa* positive primordial germ cells (Fan et al., 2008) and both *pax2a* positive CiA interneurons and *elav13* positive neurons (Cerda et al., 2009). Thus zebrafish transgenesis and FACS are two techniques that complement each other.

3.1.3 Microarray Technology

While FACS sorting provides a powerful tool to enrich for a cell population of interest, it is rarely an endpoint. Sorted cells could be characterised by reverse transcription polymerase chain reaction (RT-PCR) or immunohistochemical analysis. Sorted cells could also be cultured to enable *in vitro* examination and manipulation. The entire transcriptome of a specific cell population or cell type is of great interest and can be elucidated through

the use of expression microarrays. Indeed FACS has been used successfully to isolate a cell population of interest for microarray analysis (Reeves and Posakony, 2005, Covassin et al., 2006, Cerda et al., 2009). Expression microarrays consist of a solid substrate on which a large number of DNA sequences (probes) are placed in a defined order (Stoughton, 2005). Arrays enable the expression level of thousands of genes to be measured simultaneously, that is, the entire transcriptome of an organism can be interrogated with a biological sample of interest. This has the potential to generate vast amounts of data and to identify a role for genes never before associated with the sample being tested.

In this project gene expression levels are being compared between two different biological conditions; GFP positive cells from *sox10*:GFP transgenic lines from WT embryos versus GFP positive cells from *sox10*$^{-/-}$ embryos utilising the Affymetrix zebrafish GeneChip® platform. The zebrafish GeneChip® is made up of probes (25-mer oligonucleotides) that are organized into a square containing thousands of perfect match (PM) probes of a single and unique sequence, these squares are probe cells. Every probe cell has a corresponding mis-match (MM) probe cell consisting of the identical 25-mer oligonucleotide with a mis-match base in the middle. Each transcript on the array is recognized by 16 different probe cells, known as a probe set. The zebrafish GeneChip® has 15,509 probe sets on it and these probe sets detect approximately 14,900 different transcripts (Affymetrix, 2008).

To generate material for hybridization to an array, RNA isolated from the sample of interest is used to generate cDNA by reverse transcription using oligo-dT primers. This is followed, if required, by an amplification step to increase the amount of cDNA. The resulting cDNA is then fragmented and labelled with biotin; this generates the target material which is hybridized to the array. After hybridization the array is washed leaving biotin labelled target cDNA specifically bound to probes on the array. After washing, the array is stained using Streptavidin-phycoerythin, a fluorescent dye that binds biotin and is then scanned with a laser to detect fluorescence. This is recorded, processed and is utilised to produce a readout measuring transcript abundance (signal) for each probe cell. The signal readings are then summarised, this combines data from all the probe cells in a set to give a measure of transcript levels. Subsequently microarray data can be analysed using specialist software to compare the two groups and identify differential gene expression.

Affymetrix arrays differ from the two colour array format, where two different fluorescently labelled samples are hybridized to one array for comparison. With the Affymetrix system, each independent sample is hybridized to an array and then comparisons are made between arrays (WT arrays versus mutant arrays for example).

This requires the Affymetrix arrays to be normalized for noise and background as part of the analysis process (Harrington et al., 2000).

Microarray technology has been applied in a wide variety of organisms and contexts to elucidate biological information. The aim of this experiment was to identify genes whose expression has been perturbed in *sox10* mutant embryos to aid the discovery of genes important in the development of the NC and otic vesicle. This application of microarray technology has been successfully utilised to identify genes affected in zebrafish mutant embryos such as the *cloche* mutant (Qian et al., 2005, Sumanas et al., 2005). However both of these studies used RNA extracted from whole embryos; this has the potential to mask real changes in gene expression in the tissue of interest behind signal detected from unaffected tissues or cell types. A study of *young* mutant embryos (Leung et al., 2008) circumvented this problem by isolating the tissue of interest, the retina. There are currently no published cases of FACS being used to isolate mutant and WT zebrafish cells for examination by microarray analysis.

Microarray analysis is a tool that has already been applied to investigate key genes in the development of the neural crest. Melanocytes transfected with various isoforms of PAX3 were interrogated using microarrays to identify downstream target genes (Wang et al., 2007). Novel targets of MITF have been identified through a combination of various microarray analyses from melanomas and melanocyte cultures (Hoek et al., 2008). These studies have identified genes involved in the development of only one neural crest cell type. A macroarray approach was adopted to identify the gene expression profile of NCCs induced from chick neural tissue in an *in vitro* setting (Adams et al., 2008, Gammill and Bronner-Fraser, 2002). A large number of microarray experiments are performed on cultured tissue as it is easy to generate the tissue or cell type of interest in the quantities required for microarray analysis. Often these cultures are subjected to additional manipulations such as transfection of a gene to induce over expression or siRNA treatment to knock expression down. Thus these experiments may not recapitulate the gene expression profile present in an *in vivo* context. This problem should be resolved by isolating cells from live embryos to generate gene expression data that is as faithful to the situation in developing *in vivo* cells as possible.

Microarray studies have the potential to elucidate a genome wide gene expression profile and reveal unexpected results. However the success of this type of experiment is dependent on the quality of biological material prepared for microarray analysis. Samples need to be independent biological replicates, which are for example, carefully stage matched to limit the biological variation of factors that are not being tested. The preparation of material for array hybridization, especially, if an amplification step is included, can also introduce biases that are independent of the variable being examined.

If noise generated by biological and technological variations is too great then detecting real changes in gene expression levels could be problematic. It is also a concern that a sufficient number of arrays are compared to produce statistically significant results; this is often offset against the high cost of array experiments. The statistical analysis of microarray data will generate false positives and it is important to correctly control for these to minimise their presence. Due to the identification of false positives, candidate genes revealed by microarray data usually require validation by an independent technique.

3.1.4 Introduction to Key Marker Genes down-regulated in *sox10* mutants.

The transcription factor *sox10* is expressed in non-skeletogenic neural crest cells in the head and trunk of WT embryos (**Figure 3** and **Figure 6**). This includes, at 24 hpf, expression in pre-migratory NCCs, pigment cells migrating in the head and on the medial and lateral pathways and expression in the neurons and glia of the PNS migrating on the medial pathway and the cranial ganglia (Dutton et al., 2001a). In *sox10*$^{-/-}$ embryos the expression pattern of *sox10* is altered (**Figure 6**). There is a reduction in the number of migrating *sox10* positive cells on both pathways (Dutton et al., 2001a). Essentially pigment cells fail to specify correctly and do not migrate. As a result of this, the number of *sox10* positive pre-migratory cells is markedly increased. Another noticeable difference is the clumping of *sox10* positive cells in the head in mutant embryos, in WT embryos these cells are more evenly spread over a greater area of the head (Dutton et al., 2001a). There is also a reduction in the number of *sox10* positive cranial ganglia cells. Expression in the otic epithelium is strong in WT embryos and does not appear to be affected in *sox10*$^{-/-}$ embryos. It should be noted that the *sox10*^{m618} allele generates a full length transcript but it is essentially non-functional as a result of an amino acid substitution. To summarise, the expression pattern of *sox10* is altered in *sox10* mutants but the number of cells and transcript levels of *sox10* positive cells appears to be unchanged at 24 hpf.

There are a number of well characterised NC expressed genes that are known or expected to be down-regulated in *sox10*$^{-/-}$ embryos. The key melanocyte transcription factor *mitfa* is expressed in pre-migratory and migrating melanoblasts at 24 hpf in WT embryos. Expression of *mitfa* is very strongly reduced in *sox10*$^{-/-}$ embryos (**Figure 6**). Similarly, *dct* is strongly expressed in some melanoblasts at 24 hpf and expression is strongly reduced although not absent in NCCs of *sox10* mutant embryos (**Figure 6**). The expression of *dct* seen in the retinal pigmented epithelium is unaffected in *sox10* mutants. The xanthoblast and xanthophore marker *xdh* (also known as *aox3*) is expressed in pre-migratory and lateral pathway migrating NCCs (**Figure 6**). The expression of *xdh* is reduced in *sox10*$^{-/-}$ embryos but some pre-migratory cells and some clumped cells in the

head remain (**Figure 6**). The expression pattern of another xanthoblast and xanthophore marker, *gch2*, is very similar to *xdh* in WT and *sox10*$^{-/-}$ embryos (data not shown). The marker *gch2* should be used with caution though as it is also expressed in melanoblasts (Pelletier et al., 2001).

Crestin is widely used as a NC marker and is expressed in skeletogenic and non-skeletogenic NCCs including pre-migratory NC, NCCs migrating on the medial pathway and NCCs in the branchial arch streams (Elworthy et al., 2005). There are considerable defects in *crestin* expression in *sox10*$^{-/-}$ embryos with non-skeletogenic derivatives showing strongly reduced expression (Elworthy et al., 2005). The transcription factor *foxd3* is a marker of early crest (Odenthal and Nüsslein-Volhard, 1998) but *foxd3* expression in this early pre-migratory crest is not perturbed in *sox10* mutant embryos (Lopes et al., 2008). *Foxd3* is also expressed in the neural crest derived PNS glia surrounding cranial ganglia and posterior lateral line neurons. There is a reduction in *foxd3* positive cells at these locations in *sox10*$^{-/-}$ embryos (Kelsh et al., 2000).

Sox10 is strongly expressed in the otic epithelium of WT embryos at 24 hpf and this expression is not noticeably altered in *sox10*$^{-/-}$ embryos (**Figure 6**). A strong marker of the otic epithelium is *otomp* (Thisse et al., 2004, Murayama et al., 2005) and this gene is strongly down-regulated in *sox10*$^{-/-}$ embryos (Kelsh R, personal communication and see Chapter 4). In summary, the above NC and ear marker genes are all down-regulated to some extent in *sox10* mutant embryos and will provide useful marker genes during the establishment and validation of the *sox10* targets microarray screen.

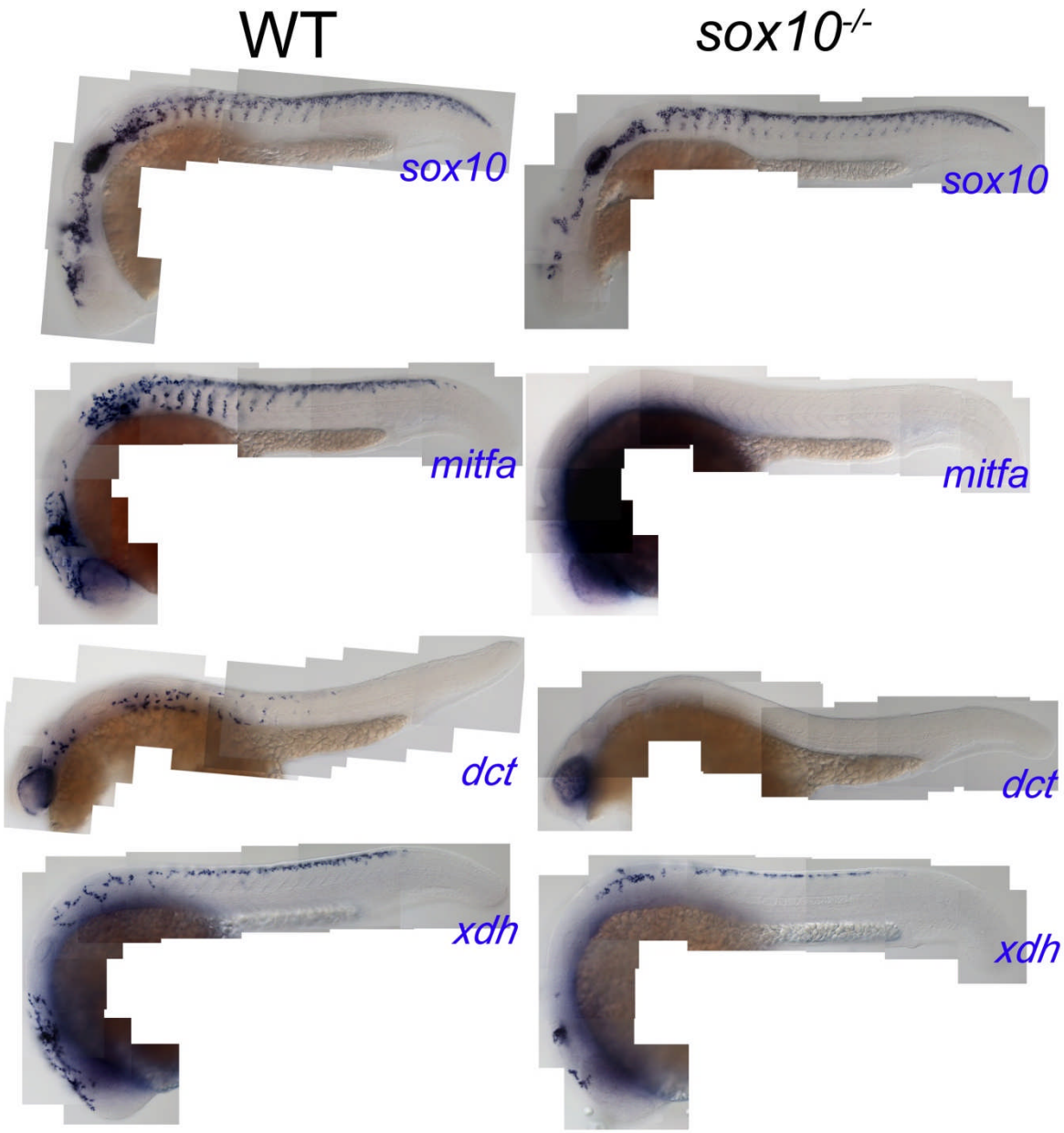


Figure 6: Whole mount *in situ* hybridizations comparing the expression patterns of four key marker genes between WT and *sox10* mutant embryos.
All images are composites of lateral views of 24 hpf embryos. The expression pattern of *sox10* is altered in mutant embryos, which display obvious clumping of *sox10* positive cells in the head and a reduced number of migrating NCCs. There is an increased density of NCCs expressing *sox10* in a dorsal position. There is no obvious difference in *sox10* expression levels in the otic epithelium of mutant embryos. Expression of *mitfa* and *dct* in the melanoblasts of WT embryos is absent in *sox10*$^{-/-}$ embryos although *dct* is still expressed in cells of the retinal pigmented epithelium. *Xdh* is expressed in pre-migratory NC and cells on the lateral pathway; this gene is a marker of xanthoblasts. Expression of *xdh* is reduced in mutant embryos and remaining *xdh* positive cells show an altered distribution. *Xdh* positive cells are clumped in the head and stuck in pre-migratory positions in mutant embryos.

3.1.5 Aims

The ultimate goal of this section of work was to generate microarray data from 24 hpf WT and *sox10*$^{-/-}$ zebrafish embryos for comparison. This was to enable the identification of genes differentially expressed in mutant embryos. The 24 hpf stage was chosen because *sox10* is extensively expressed in the NC and otic epithelium at this time point, but more importantly NCCs are undergoing and expressing key genes for specification, differentiation and migration. The goal of this project is to identify genes that are important for these processes in the gene regulatory network of *sox10*. To achieve this, the following aims had to be fulfilled:

- A method had to be established to identify *sox10*$^{-/-}$ from WT embryos at a stage prior to 24 hpf, preferably as early as possible.
- A reliable protocol to enrich for GFP positive cells of interest from transgenic zebrafish embryos by FACS had to be established.
- RNA had to be extracted from these groups of cells that was of suitable quality and quantity for Microarray analysis on 10 chips (5 mutant and 5 WT) at the DNA and Protein Microarray Facility of the University of California, Irvine.
- Analysis of the Microarray data to generate a list of differentially regulated targets for further validation had to be performed.

3.2 Results and Conclusions

3.2.1 Identification of *sox10* mutant embryos during early development

To enable the sorting of both WT and *sox10*$^{-/-}$ NCCs for microarray analysis, transgenic embryos from both these genotypes had to be identified to a high degree of accuracy at an early stage, around 20 hpf or late segmentation stage. With the aim being to generate a gene expression profile of cells at approximately 24 hpf, embryos had to be identified at least two hours prior to this time point to ensure that embryos could be disaggregated and the resulting single cells processed by FACS. Zebrafish embryo pigmentation (melanisation) is first visible at 28 hpf and provides an excellent visual marker for genotyping WT and *sox10* mutant embryos but these embryos needed to be genotyped earlier in development using visible GFP reporter expression. Initially this was achieved with the *7.2sox10*:GFP line and subsequently with the *4.*9*sox10*:GFP line. It should be noted that throughout this thesis, the non-mutant transgenic embryos are referred to as WT embryos for brevity.

When examining *7.2sox10*:GFP embryos at 20-22 hpf, from a cross where 25 % of the progeny were *sox10*$^{-/-}$ it was noted that a subset of embryos displayed a clumping of GFP expressing cells in the head and reduced GFP expression in the otic vesicle. From 3

pairs of adult fish 26.5%, 30.5% and 22.8% of embryos showed clumping of GFP positive cells. It was postulated that these were *sox10*$^{-/-}$ embryos. This hypothesis was tested by allowing embryos to develop; embryos sorted by these characteristics did not develop melanophores or other types of pigment cell. Using these criteria for genotyping, embryos were sorted into WT (non-mutant) or *sox10*$^{-/-}$ groups without any contamination. Examples of GFP expression in these two genotypes at 27 and 30 hpf are shown in **Figure 7**. Arrows in **Figure 7B** and **D** mark the clumping of GFP positive cells in the head. In WT embryos the GFP positive cells are evenly distributed over the majority of the head (**Figure 7A** and **C**). Asterisks highlight the otic vesicle; note the reduced level of GFP in **Figure 7B** and **D** compared to **Figure 7A** and **C**. Additionally, clumping of cells in the heads of *sox10*$^{-/-}$ embryos at 24 hpf was also noted when performing *in situ* hybridizations for key NC marker genes such as *sox10*, *mitfa*, and *xdh* (**Figure 6**) as well as numerous genes identified during the *in situ* hybridization microarray validation screen (see Chapter 4).

Having successfully identified how to genotype embryos from the *7.2sox10*:GFP line, this exercise was repeated with the *4.*9*sox10*:GFP line. As the otic epithelium expression is much weaker in this line it did not provide a suitable marker for genotyping WT from *sox10*$^{-/-}$ embryos. The pattern of clumped GFP positive cells in the head however appeared to be the same as observed with the *7.2sox10*:GFP line and again provided a suitable marker for genotyping (Data not shown). Mutant embryos could be distinguished from WT embryos without cross contaminating the two genotypes.

Differences were identified in the pattern of GFP expression between WT and mutant embryos for both transgenic lines. WT and mutant embryos could be distinguished at stages sufficiently early (approximately 22 hpf) to enable sorting of NCCs from 24 hpf embryos, the time point of interest. Also by sorting the embryos with one hundred percent accuracy pure batches of WT and mutant embryos could be isolated that would generate GFP positive cells whose expression profile could be resolved and compared.

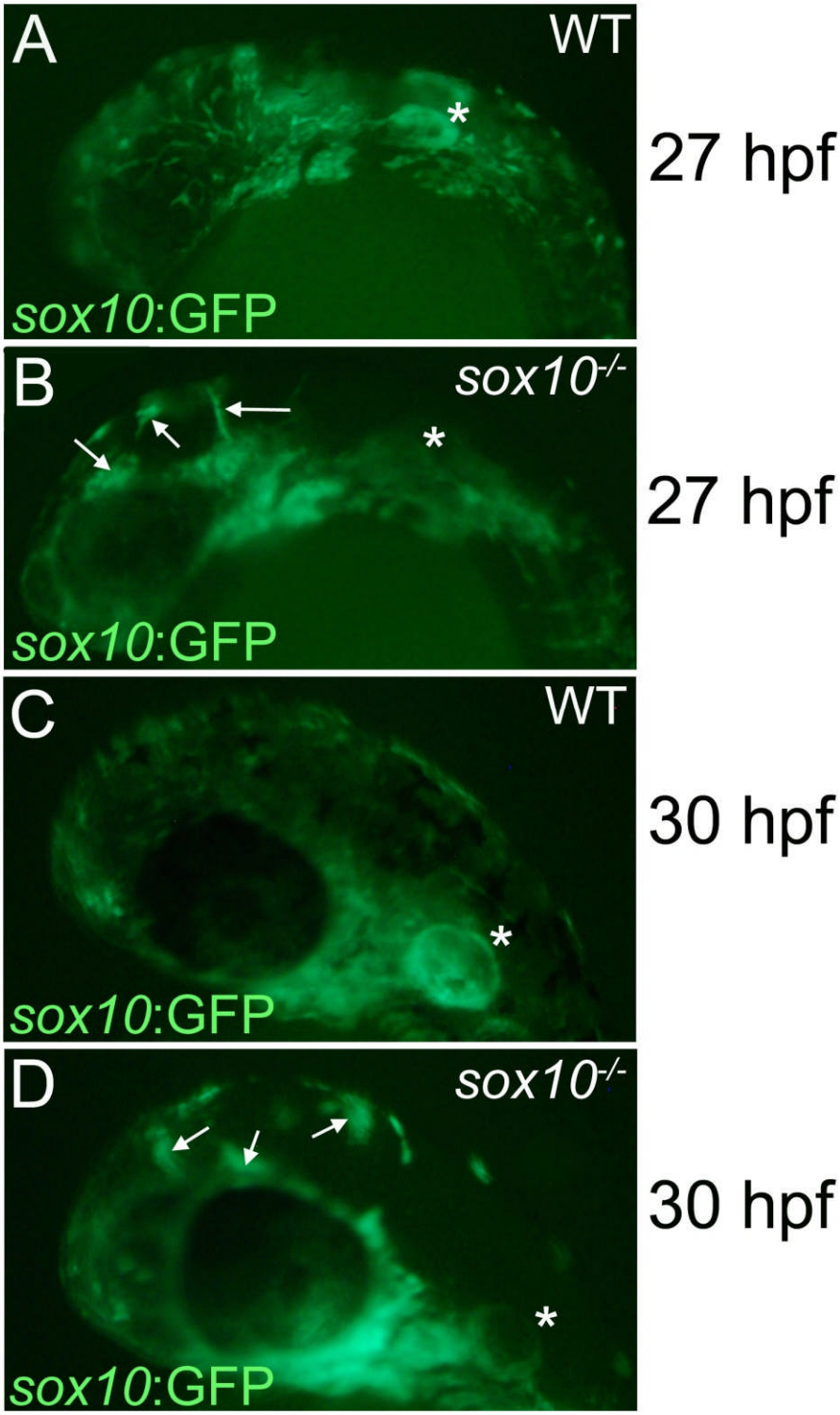


Figure 7: Genotyping WT and *sox10*$^{-/-}$ embryos during early stages of development. All images show *7.2sox10*:GFP embryos, GFP visualised under UV light on a dissecting microscope with embryos in the same orientation, dorsal top, ventral bottom, anterior left and posterior right. Embryos visualised at 27 hpf (A and B) and at 30 hpf (C and D). Note the clumping of GFP positive cells in B and D indicated by arrows that are absent in A and C where GFP positive cells are evenly distributed over the entire head of the embryo. Also note the reduced expression of GFP in the otic vesicle (*) when comparing B to A and D to C. In C melanophores can be seen starting to melanise thus confirming the embryo is WT.

3.2.2 FACS sorting of NCCs

Having genotyped WT from *sox10*$^{-/-}$ embryos at a suitably early stage, GFP positive cells from both populations of embryos (mutant and WT) could be purified by FACS. From a cross of adult zebrafish heterozygous for the *sox10*^{m618} allele and in which one parent was heterozygous for a *sox10:*GFP transgene, 500 GFP positive embryos were selected and then dechorionated using forceps. This occurred at approximately 20 hpf when GFP was clearly visible. These dechorionated embryos were lightly anaesthetised with tricaine and sorted by GFP pattern. To sort embryos for GFP and then to sort for WT and *sox10*$^{-/-}$ embryos took 3 to 4 hours thus embryos were disaggregated at approximately 24 hpf. 100 embryos of each genotype heavily anaesthetised with tricaine were placed into separate Eppendorf tubes and disaggregated with a micropestle. Cellular clumps generated by this were then broken down to single cells for FACS using a 35 minute trypsin digest at 33°C. Approximately 24 hpf embryos were disaggregated, a process which took an hour, single cells were therefore approximately 25 hpf when FACS was initiated.

At the start of each FACS session 50,000 WT cells were run to visualise GFP positive and GFP negative cells (**Figure 8**). Using this run the gates were then set to determine the GFP positive and the GFP negative cells to be collected in RLT buffer. Typically, for the mutant sample four runs of 500,000 cells were then carried out and the sorted cells were pooled together. Each run took between 8 and 10 minutes to complete and the cells were mixed with RLT buffer by pipetting. Only 2 million events were run to minimise the period of time it took to sort cells. This was then repeated to collect the GFP positive WT cells. One clutch of embryos provided a mutant sample and a WT sample for RNA extraction. No two mutant samples came from the same clutch of embryos thus making them independent biological replicates with the same true of the WT samples. Total numbers of cells sorted are shown in **Table 5** for the *7.2sox10*:GFP line and **Table 6** for the *4.9sox10*:GFP line. From the *7.2sox10*:GFP line 6 mutant samples (7.2_M1 to M6) and 6 wild type samples (7.2_WT1 to WT6) were generated. From the 4.9*sox10*:GFP line 4 mutant samples (4.9_M1 to M4) and 4 wild type samples (4.9_WT1 to WT4) were produced.

Running two million cells from the *7.2sox10*:GFP line through the FACS machine collected between 40,000 and 60,000 GFP positive cells (**Table 5**). This data shows that 2-3 % of the cells from these transgenic embryos express GFP. Sorted cells are predicted to include GFP positive NCCs and cells from the otic epithelium (**Figure 7**). The number of cells collected appears to be lower from WT embryos than from mutant embryos. The mean number of GFP positive events ± standard deviation sorted during a 500,000 event run from mutant samples equalled 13815 ± 1205 while for WT samples it equalled 11933 ± 1583. A two tailed Student's *t*-Test gave a *p*-value of less than 0.0001, where the null

hypothesis was no difference, demonstrating that there is a significant statistical difference between the mutant and WT samples.

While sorting cells during run 4 (marked by an asterisk in **Table 6**) to generate sample 4.9_WT1 the sample tubes that collect cells were removed to refocus the liquid streams mid-run. As part of this process the system is flushed with Becton Dickinson FACSRinse solution to clear any blocks. This sorting run was not paused during the flushing process and thus registered a large number of additional events that were not collected. In fact fewer cells would have been collected during this run than normal as the 500,000 event limit would have been reached with a short period during which nothing was collected. This data set has been ignored in the following analyses but is shown in the table to be comprehensive.

The number of cells collected by FACS from the *4.9sox10*:GFP line (**Table 6**) from a 500,000 event run appeared to be less than sorted previously from the *7.2sox10*:GFP line (**Table 5**). This was noted after gathering the M1, M2, WT1 and WT2 samples and the difference was significant (two tailed Student's *t*-Test $p<0.0001$). Cells from an additional 500,000 event run (five runs or 2.5 million events each in total) were collected to generate samples M3, M4, WT3 and WT4 in an attempt to compensate for this. The mean number of GFP positive cells sorted from the *7.2sox10*:GFP line ± standard deviation equalled 12874 ± 1686 and for the 4.9*sox10*:GFP line equalled 9828 ± 2386. Between 30,000 and 60,000 GFP positive cells were sorted to make up each *4.9sox10*:GFP sample. For most samples, 1.5 – 2 % of cells from this transgenic line were GFP positive, again this seemed to be a lower percentage than collected from the *7.2sox10*:GFP line and tallied with the *t*-Test result. The *4.9sox10*:GFP line does not express GFP strongly in the otic epithelium therefore these cells may not be selected by FACS. As seen when sorting cells from the *7.2sox10*:GFP line, it appeared that the WT samples generated fewer GFP positive cells than the mutant samples. This was tested using a two tailed Student's *t*-Test (*p*-value = 0.054), therefore the null hypothesis of no difference could not be rejected. The mean number of GFP positive cells collected from a 500,000 event run ± standard deviation was 10578 ± 2273 for mutant samples and 8530± 2302 for wild type samples.

In conclusion, GFP positive cells from two transgenic lines have been sorted using FACS on WT and *sox10*$^{-/-}$ backgrounds. As such, a small population of cells from disaggregated whole transgenic embryos have been enriched, a population of cells that accounts for only 1.5 - 3% of an embryo.

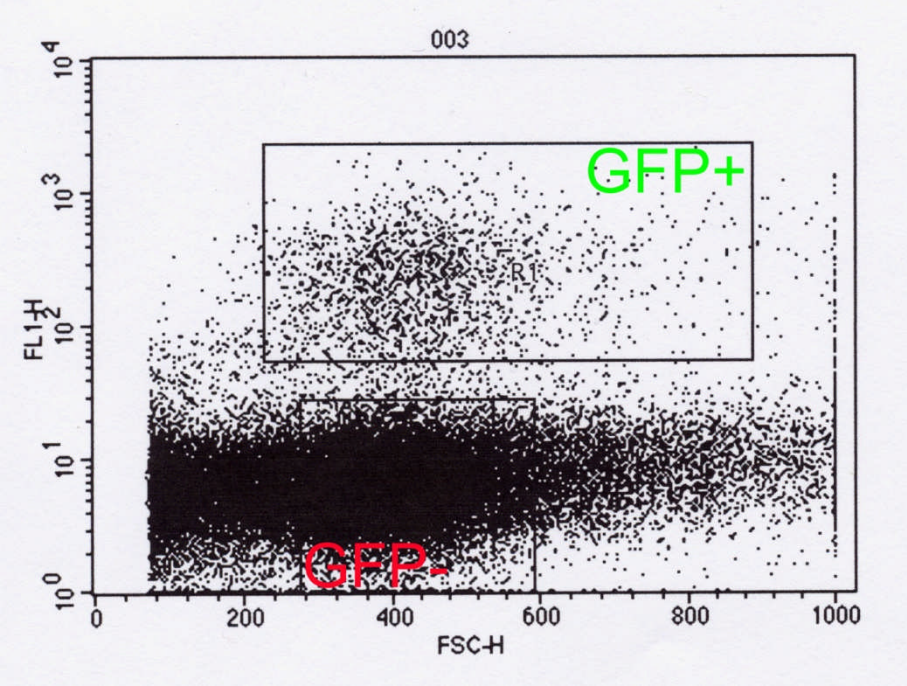


Figure 8: Dot plot visualizing a 50,000 cell FACS run.
This dot plot shows 50,000 events (single cells) recorded during a FACS run. The gates set for collecting cells using this data are also shown by boxes. The upper gate defines the GFP positive population of cells sorted while the lower gate marks the GFP negative cells collected. Note that the gates are set around obvious concentrations of events. The x-axis shows FSC or forward scatter, this is roughly proportional to cell size. The y-axis displays detected fluorescence.

Run Number	7.2_M1	7.2_M2	7.2_M3	7.2_M4	7.2_M5	7.2_M6
1	14061	14919	13614	15419	11245	14817
2	13790	14018	14466	15106	11862	14984
3	13840	13735	14028	14801	11203	14349
4	13374	12996	14211	14214	11731	14782
Total	55065 (2.75 %)	55668 (2.78 %)	56319 (2.82 %)	59540 (2.98 %)	46041 (2.30 %)	58932 (2.95 %)
Run Number	7.2_WT1	7.2_WT2	7.2_WT3	7.2_WT4	7.2_WT5	7.2_WT6
1	12577	14027	10759	14691	10342	13981
2	9808	12816	10641	14265	10721	12822
3	9714	12443	10853	13315	10880	12842
4	9818	10760	10973	13366	10677	13309
Total	41917 (2.10 %)	50046 (2.50 %)	43226 (2.16 %)	55637 (2.78 %)	42620 (2.13 %)	52954 (2.65 %)

Table 5: Numbers of events sorted from 7.2*sox10*:GFP embryos.
The number of GFP positive events (cells) sorted from individual 500,000 event runs are shown. Runs that were pooled together are in columns and total numbers of cells pooled are shown. The percentage shown next to total counts is the percentage of cells sorted as GFP positive from all events detected. The mean ± standard deviation for all mutant runs = 13815±1205 and for all WT runs = 11933±1583.

Run Number	4.9_M1	4.9_M2	4.9_M3	4.9_M4
1	14469	9521	10164	9566
2	14475	9590	9605	9114
3	15050	9331	10047	8970
4	14553	9323	10052	8677
5	-	-	9245	8653
Total	58547 (2.93 %)	37765 (1.89 %)	49113 (1.96 %)	44980 (1.80 %)
Run Number	4.9_WT1	4.9_WT2	4.9_WT3	4.9_WT4
1	12796	8980	9677	9136
2	12832	8225	7777	8946
3	13490	8456	8359	8388
4	57740*	8010	8395	8142
5	-	-	3570	8402
Total	96853 (4.84 %)	33671 (1.68 %)	37778 (1.51%)	43014 (1.72 %)

Table 6: Number of events sorted from 4.9*sox10*:GFP embryos.
The number of GFP positive events (cells) sorted from individual 500,000 event runs are shown. Presented as per Table 5. For sample WT1 a counting error has erroneously generated a very high number of sorted GFP positive cells (*) (see text).The mean ± standard deviation for all mutant runs = 10578± 2274 and for all WT runs = 8530± 2302.

3.2.3 RT-PCR analysis of sorted GFP positive cells

To determine if key NC and otic epithelium marker genes were being enriched in GFP positive samples when compared to GFP negative cells, an RT-PCR approach was adopted. Total RNA was extracted from the sorted cells using the Qiagen RNeasy® Micro kit and used to generate cDNA. The cDNA samples were amplified by polymerase chain reaction (PCR) using primers that target key marker genes (**Table 3**). The same PCR program (see Chapter 2), with a 58 °C annealing step, was used to amplify cDNA from both the *7.2sox10*:GFP and the *4.9 sox10*:GFP lines using the same primers. The *otomp* primers were used to amplify cDNA in a separate PCR reaction identical but for an annealing temperature of 48°C. DNA samples amplified by PCR were visualised by gel electrophoresis (**Figure 9**). Amplified *actin* bands from both lines appear to be slightly more intense in GFP positive samples thus the differences in the key marker gene band intensities should be viewed in this context. Despite this, a clear difference is visible for all marker genes, with stronger band intensities visible in GFP positive samples when compared to GFP negative samples. This holds true for sorted cells from both transgenic lines. The GFP positive and negative *dct* bands displayed the least striking difference in intensity. The reason for this is likely to be that *dct* is also expressed in cells of the retinal pigmented epithelium and is thus not restricted solely to the NC. Surprisingly GFP and *sox10* was amplified from GFP negative samples suggesting that there was some incorrect sorting leading to GFP positive cells contaminating the GFP negative sample. In

conclusion, FACS has enabled the strong enrichment of the cell types of interest from whole zebrafish embryos.

To assess if the sorted GFP positive cells from WT and *sox10*$^{-/-}$ displayed differential gene expression an RT-PCR approach was employed as described above. DNA samples amplified by PCR were visualised by gel electrophoresis (**Figure 10** and **Figure 11**). Amplified *sox10* and *actin* cDNA maintained relatively consistent levels of expression across all samples. As expected we observed that *dct* and *mitfa* had reduced transcript levels in the mutant samples compared to the WT samples. This demonstrated that the expression of two key NC genes involved in the development of melanophores and known to be affected in *sox10* mutants were reduced (Kelsh and Eisen, 2000) This result matches the reduced expression of both *dct* and *mitfa* seen by *in situ* hybridization in *sox10*$^{-/-}$ embryos compared to WT siblings. The confirmation of differential gene expression between WT and mutant samples demonstrated that WT and mutant genotypes had been successfully separated. It was concluded that the total RNA extracted from the FACS samples could be sent for microarray analysis.

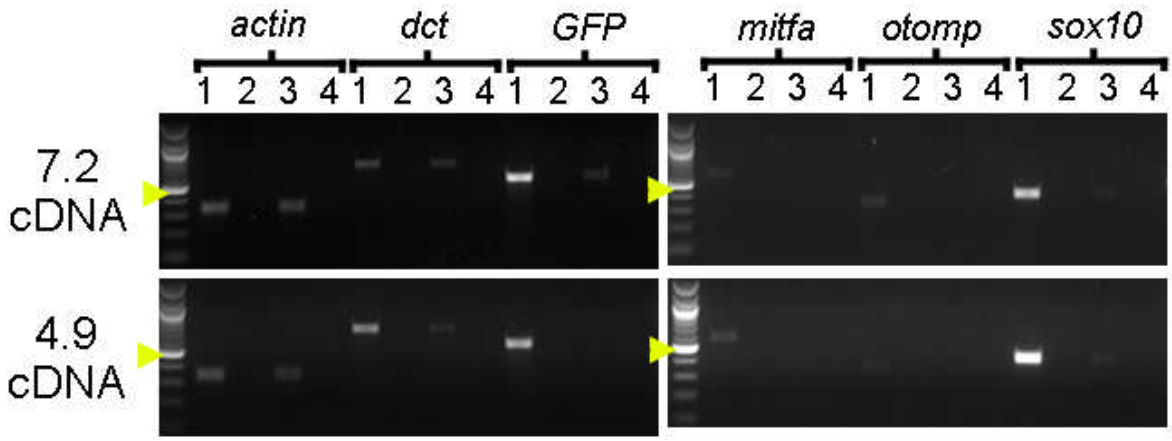


Figure 9: RT-PCR Validation of Key Marker Gene Enrichment in GFP positive cells. GFP positive and GFP negative cells were sorted by FACS from both the *7.2sox10*:GFP (top row of gels) and *4.9sox10*:GFP (bottom row of gels) lines. RNA was extracted from these samples and used to generate cDNA. This cDNA was amplified using primers that target key NC marker genes (*dct, mitfa* and *sox10*), a marker of the otic epithelium (*otomp*) and GFP. Actin primers were used as a positive control and samples lacking cDNA were used as a negative control. Amplified DNA was visualised by gel electrophoresis. A 100 bp ladder (Promega) was used as reference for band sizes; the 500 bp band is highlighted (yellow arrowhead). Lanes 1-4 for each primer pair were identical, containing samples as follows, 1 = GFP+ cDNA+, 2 = GFP+ cDNA-, 3 = GFP- CDNA+ and 4 = GFP- cDNA-. The expected band sizes amplified from cDNA were as follows: *dct* 697 bp, *mitfa* 581 bp, *sox10* 420 bp, *otomp* 355 bp, GFP 600bp and *actin* 330 bp. All bands appear to be of the expected size. Actin was amplified from the GFP positive and GFP negative samples of both lines, the band intensity of GFP positive samples was slightly greater suggesting that the amount of cDNA starting material was slightly overestimated. Band intensity from GFP positive cDNA was considerably stronger than from GFP negative cDNA for all key marker genes. GFP and *sox10* display the most striking differences in band intensity while the difference between the *dct* bands was less distinct. No bands for *mitfa* or *otomp* were detectable when amplified from the GFP negative samples but weak bands were visible in the GFP positive samples.

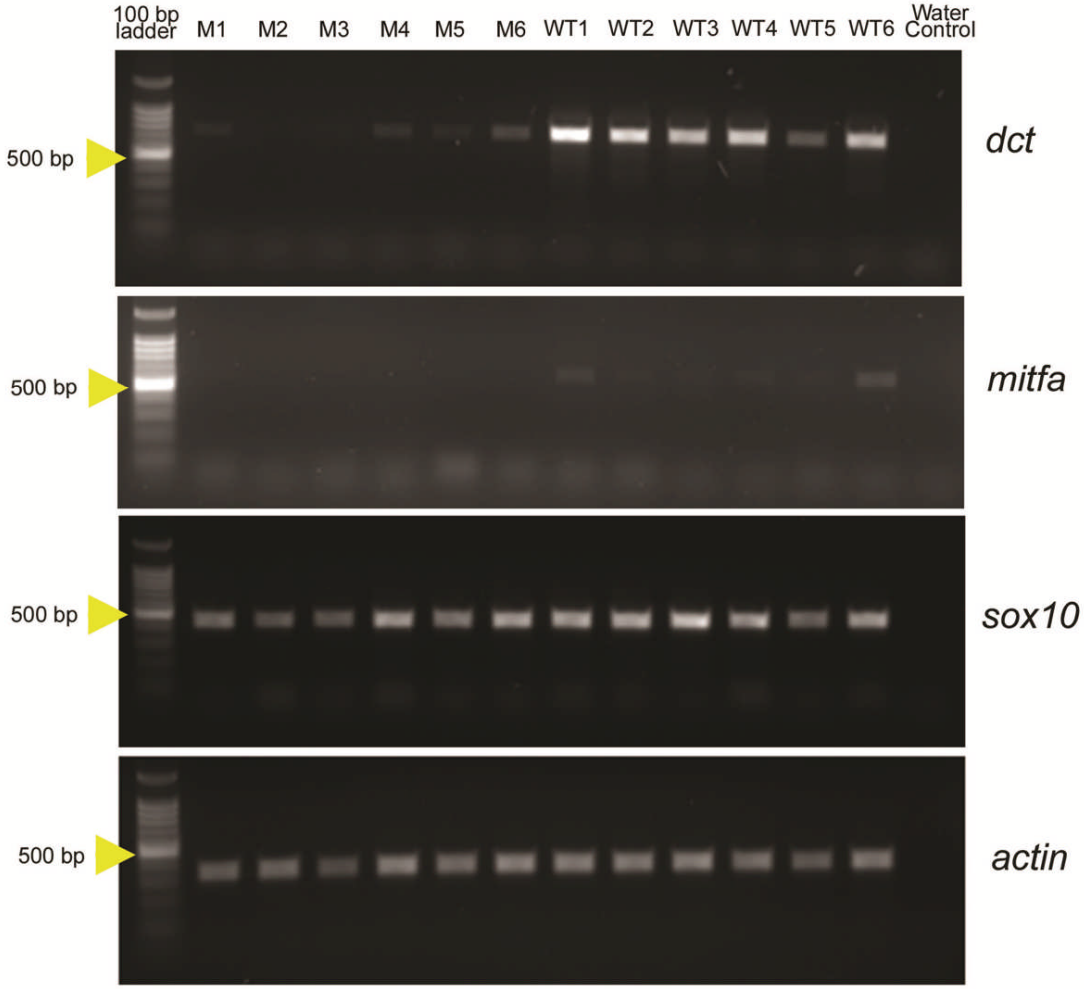


Figure 10: RT-PCR validation of differential gene expression in *7.2sox10*:GFP RNA samples.
Key NC marker genes were amplified from mutant and WT template cDNA generated using RNA extracted from FACS samples. The identifier of each sample matches that shown previously with the FACS data. *Actin* primers were used as a positive control and to show consistent amplification across all samples. A water control lacking cDNA was performed as a negative control. A 100 bp ladder (Promega) was loaded on the gel for referencing band sizes. The expected band sizes amplified from cDNA were as follows: *dct* 697 bp, *mitfa* 581 bp, *sox10* 420 bp and *actin* 330 bp. All bands appear to be of the expected size. If genomic DNA is amplified with the *actin* primers a 680 bp band is generated, this is not detected here. This provides a simple method to detect for genomic DNA contamination. Amplification of *actin* and *sox10* was reasonably consistent across all samples. *Mitfa* was not detected in any of the mutant samples but weak bands were present in all WT samples. *Dct* bands from WT samples showed very strong intensity but were much weaker in mutant samples.

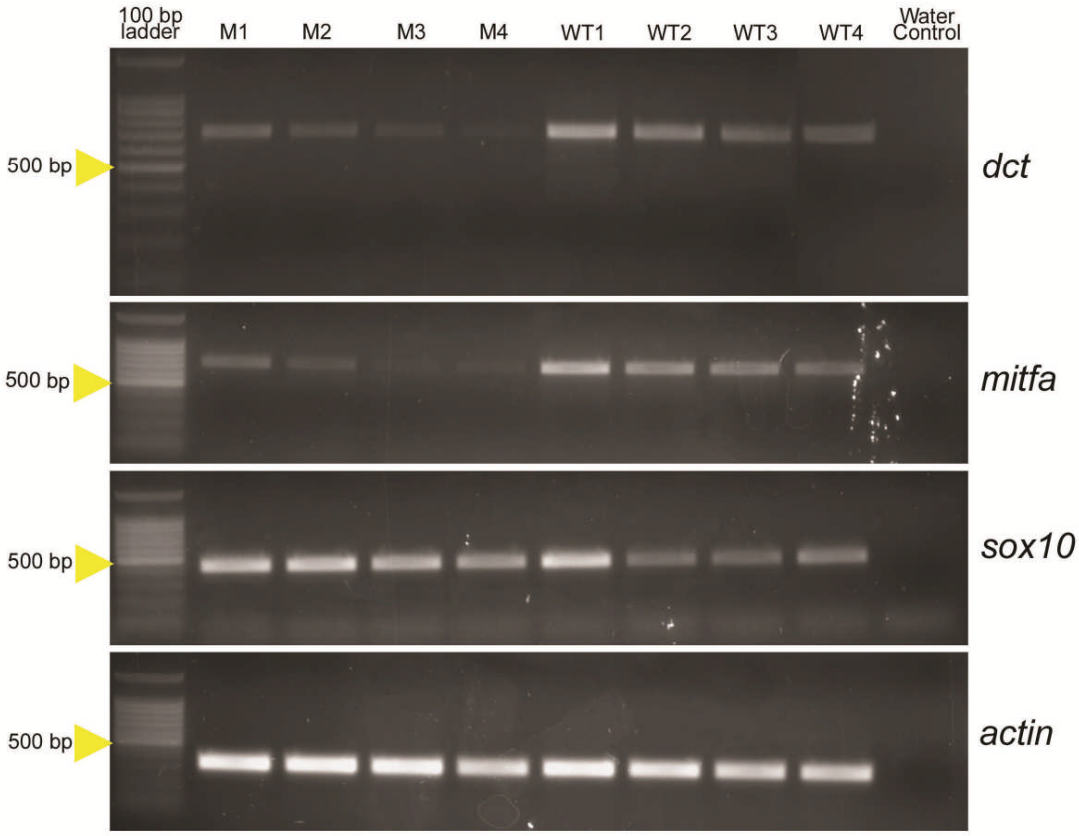


Figure 11: RT-PCR validation of differential gene expression in *4.9sox10*:GFP RNA samples.
Key NC marker genes were amplified using mutant and WT template cDNA generated from RNA extracted from FACS samples. Presented as per Figure 10.

3.2.4 Analysis of Extracted RNA

The RNA extracted from FACS samples was sent on dry ice to the UCI DNA and Protein Microarray Facility at the University of California, Irvine. Here they initially analysed RNA quality using an Agilent 2100 Bioanalyser and quantity of RNA using a Nanodrop Spectrophotometer to assess the suitability of samples for microarray analysis. The data from these analyses are shown here for completeness but demonstrate that some samples were of sufficient quality and quantity to undergo processing for microarray analysis.

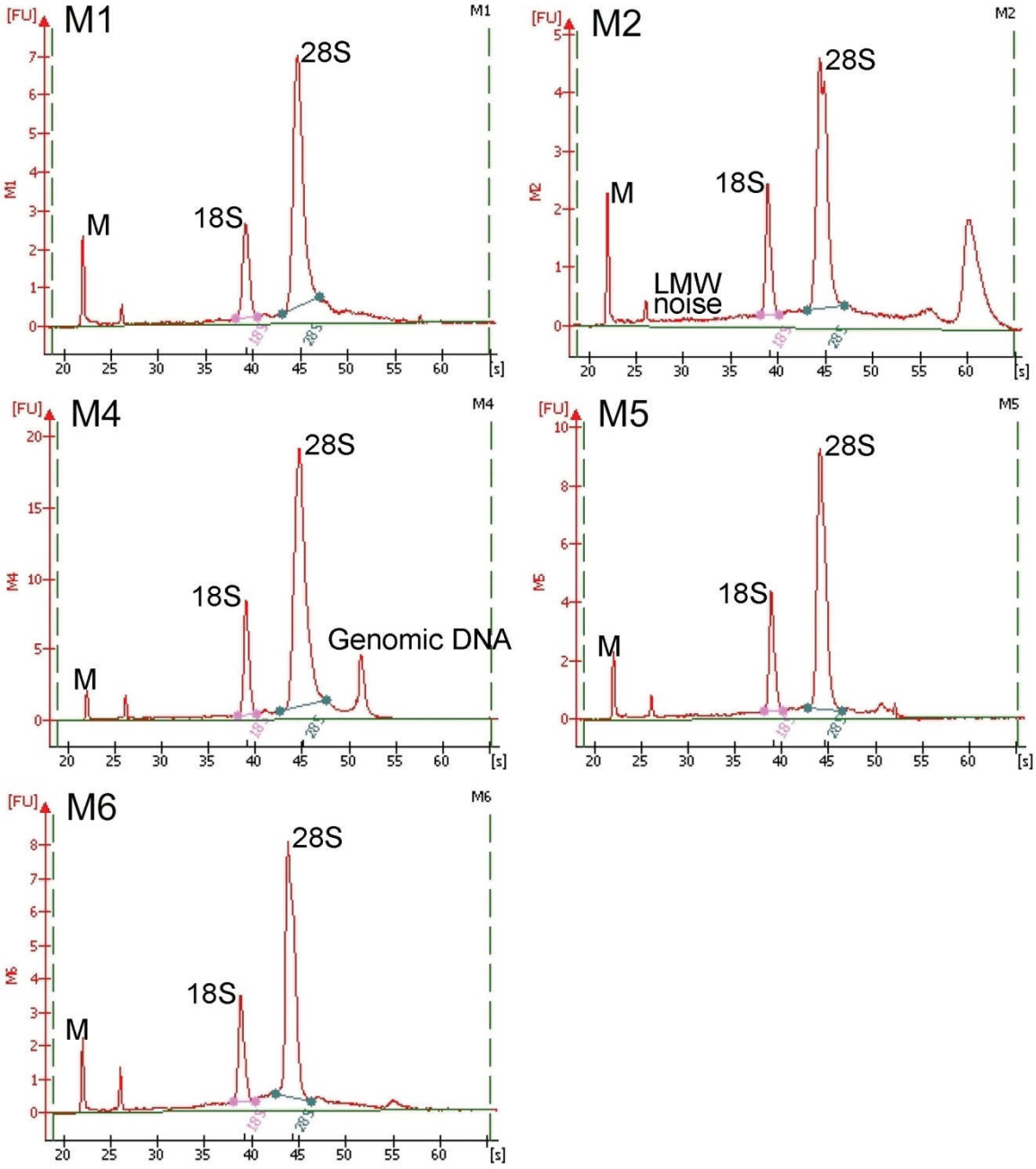
The quantity of RNA that had been extracted from the *7.2sox10*:GFP samples was very small, between 15 to 60 ng (**Table 7**), so an amplification step was required to generate enough cDNA for subsequent labelling and hybridization to the arrays. The UCI Microarray Facility had tested the WT - Ovation™ Pico RNA Amplification System from NuGEN and found it capable of generating enough cDNA for hybridization to Affymetrix GeneChips from 10 ng of RNA (C. Selby, personal communication). As such all of the samples contained enough RNA for amplification, the decision as to which samples to use was therefore based on the quality of RNA extracted. Analysis of the electropherograms

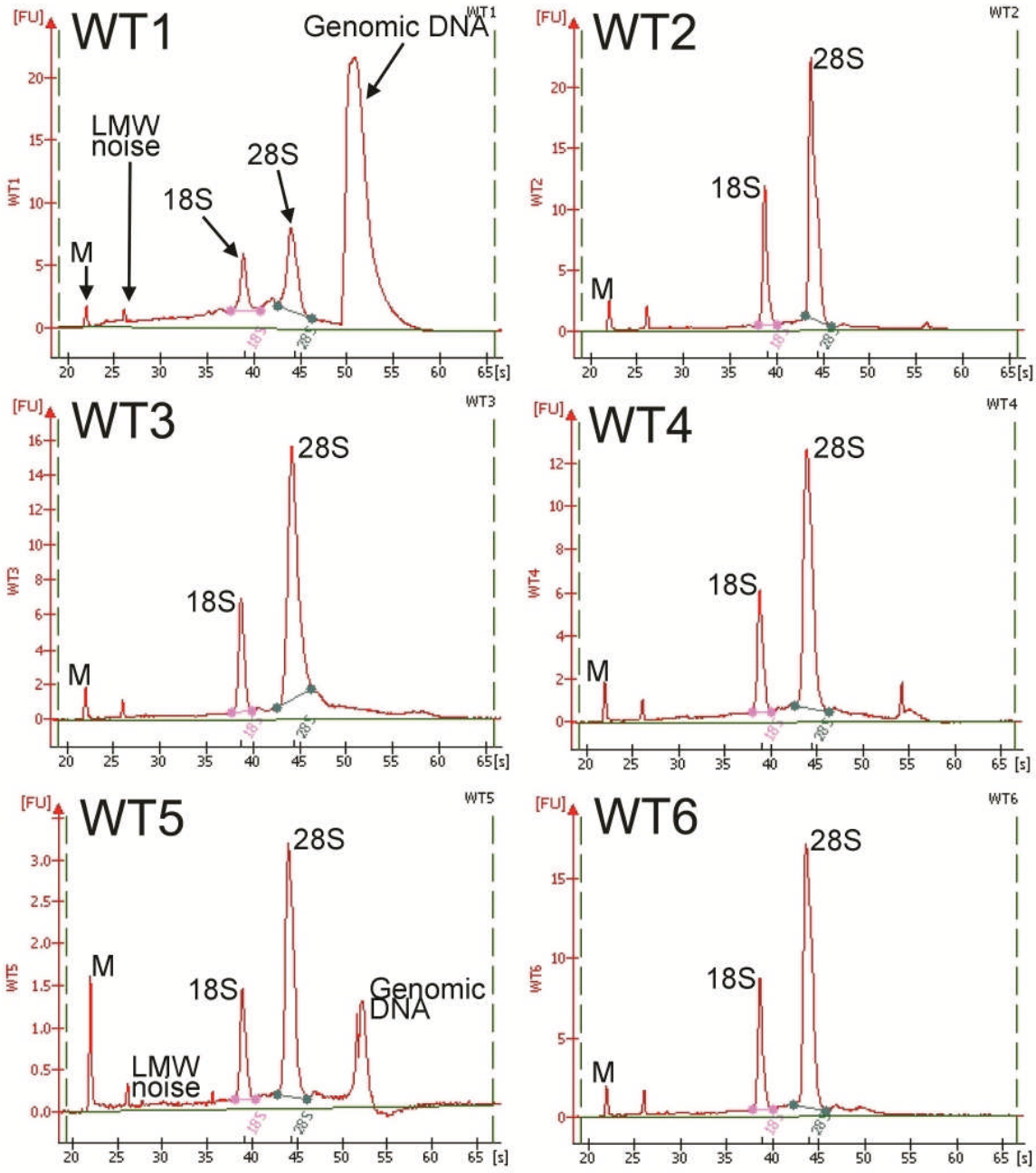
generated by the Agilent 2100 Bioanalyser was used to assess RNA quality (**Figure 12** and **Figure 13**). This assay uses fluorescence to detect the RNA sample and measures this against time. Small molecules such as degraded RNA appear quickly on the left of the electropherogram while large molecules such as genomic DNA appear to the right. A typical electropherogram of a high quality RNA sample shows two clear peaks of fluorescence corresponding to the 28s and 18s ribosomal RNA subunits, low noise between peaks and minimal low molecular weight contamination. All of the *7.2sox10*:GFP RNA samples show a good RNA profile however samples M2, M4, WT1 and WT5 show a large peak on the far right (particularly clear in sample WT1), representing genomic DNA contamination (**Figure 12**). The electropherogram for M3 is not shown but also displayed strong genomic DNA contamination. Based on this, samples M1, M5, M6, WT2, WT4 and WT6 were selected for further processing for microarray analysis.

Sample (1:3 dilution)	A260	ng/µl undiluted	Total ng
M1	0.019	2.28	20.52
M2	0.015	1.83	16.47
M3	0.013	1.62	14.58
M4	0.058	6.93	62.37
M5	0.033	3.96	35.64
M6	0.047	5.64	50.76
WT1	0.036	4.32	38.88
WT2	0.037	4.44	39.96
WT3	0.022	2.64	23.76
WT4	0.033	3.96	35.64
WT5	0.018	2.13	19.17
WT6	0.043	5.19	46.71

Table 7: Nanodrop Spectrophotometer analysis of the *7.2sox10*:GFP RNA samples.

Figure 12: Electropherograms for assessing the quality of the *7.2sox10*:GFP samples.
Electropherograms display time on the *x* axis and fluorescence on the *y* axis. Molecules are eluted at different times dependent on size with small molecules eluting first and large molecules later. RNA 18S and 28S peaks as well as genomic DNA peaks are labelled. LMW noise = Low molecular weight noise, M = Size Marker. (See next page).


M1
[FU]
M
18S
28S
[s]
M2
LMW
noise
M4
Genomic DNA
M5
M6



RNA extracted from the *4.9sox10*:GFP samples was also assessed for quantity (data not shown) and quality (**Figure 13**). The M1 electropherogram shows a profile of badly degraded RNA while that of M2 shows the profile of a marginally degraded RNA sample. The RNA profiles for both M3 and M4 appeared to be good. For the WT samples, WT3 and WT4 appeared to have the highest quality RNA with distinct single peaks corresponding to 18s and 28s ribosomal RNA. Consequently the samples M3, M4, WT3 and WT4 were chosen for further processing.

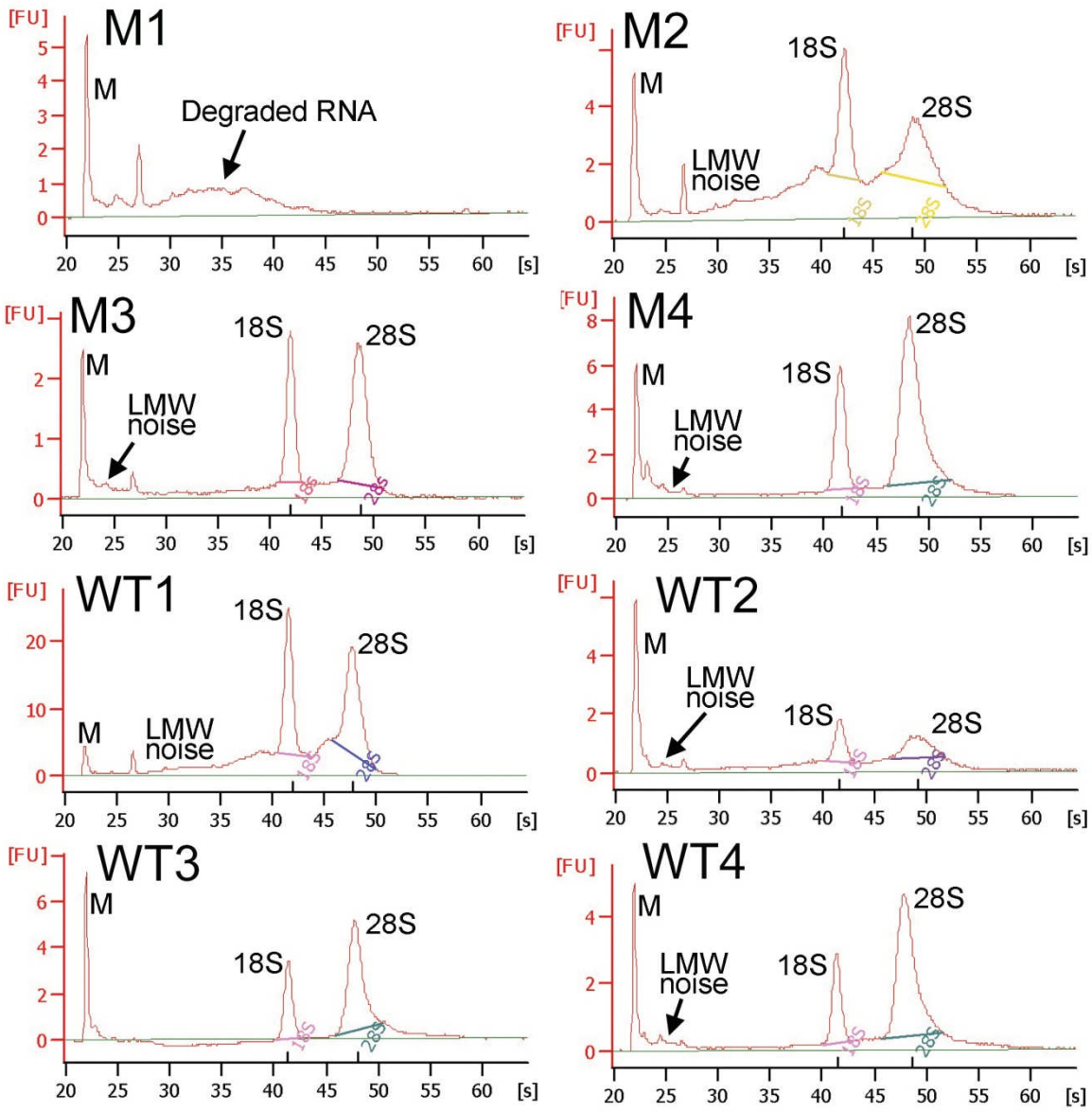


Figure 13: Electropherograms for assessing the quality of the *4.9sox10*:GFP samples.
Electropherograms display time on the *x* axis and fluorescence on the *y* axis. Molecules are eluted at different times dependent on size with small molecules eluting early and large molecules later. RNA 18S and 28S peaks and genomic DNA peaks are labelled. LMW noise = Low molecular weight noise, M = Size Marker.

The best quality RNA samples were selected to be processed for hybridization to the microarrays. The samples were used as a template to generate cDNA followed by an amplification step. The cDNA was then fragmented, labelled with biotin and hybridized to the Affymetrix zebrafish GeneChips. Initially the UCI DNA Microarray Facility was asked to process two samples, M6 and WT6 for a first analysis. Once the results from this were determined and appeared to be promising, the remaining *7.2sox10*:GFP samples were processed. The *4.9sox10*:GFP samples were generated after the *7.2sox10*:GFP samples and were processed subsequently in a single batch.

3.2.5 Initial Analysis of Microarray Data

Initial assessment and analysis of the array data provided by the UCI Microarray facility was performed with the Affymetrix GeneChip® Operating Software (GCOS) using guidance presented in the Affymetrix Data Analysis Fundamentals handbook (Affymetrix, 2008). The data from the *7.2sox10*:GFP line consisted of three WT versus three mutant arrays, the data from the *4.9sox10*:GFP line consisted of two WT versus two mutant arrays.

3.2.5.1 Assessing GeneChip® Quality Control Data

The array data received from the UCI Microarray Facility was assessed using the quality control guidelines specified in the Affymetrix Data Analysis Fundamentals handbook.

Poly – A RNA controls, *lys, phe, thr* and *dap* are polyadenylated prokaryotic controls that are spiked into the RNA sample mix and can be used to monitor the entire target labelling process. They serve as a positive control for target preparation and labelling, independent of the quality of the extracted RNA sent for microarray analysis. All were present on all of the scanned arrays as expected (**Table 8** and **Table 9**).

The hybridization controls are a mix of biotin labelled cRNA controls that are spiked into the hybridization mix and are used to assess the efficiency of hybridization independently of the sample preparation stages. *BioB* should be detected as present at least 50 % of the time. The remaining hybridization controls should always be present. All of the control hybridizations are present on all of the chips (**Table 8** and **Table 9**).

The Internal control genes β-actin and GAPDH are used to assess RNA sample and assay quality. The signal values from the 3' probe sets of these genes are compared with the corresponding 5' probe sets to produce a ratio of 3'/5'. Typically the ratio of 3' to 5' is no more than 3 but rounds of amplification can increase this (Affymetrix, 2008). A high ratio of 3'/5' can be an indication of RNA degradation or inefficiencies in the *in vitro* transcription reaction that generates double stranded cDNA. The 3'/5' ratio for GAPDH on all of the arrays was between 1 and 1.55. The 3'/5' ratio was consistently much higher than 3 on all of the arrays for actin (**Table 8** and **Table 9**).

The number of probes identified as present, based on signal values, relative to the total number of probe sets on the entire array is represented as a percentage. This value is dependent on the biological sample being tested but should be roughly consistent for replicate samples. A very low percentage can be an indication of poor sample quality. All of the arrays except 4.9_WT3 displayed 50 % or more of probes as present, 4.9_WT3 had only 45 % of probes present (**Table 8** and **Table 9**). Across these 10 samples, derived

from similar biological sources, the percentage of probes present ranged from 45 – 60 % but was typically similar with 4.9_WT3 a slight outlier.

Average background is reported as a signal value detected from across the entire array. Typically background should be between 20 -100 (Affymetrix, 2008). All of the arrays displayed background within this range with values of roughly 50 to 75. Arrays being compared should have similar background values. Noise (RawQ) is a measure of pixel to pixel variation of probe cells on a chip, the main contribution to noise comes from the electrical noise of the scanner. Arrays that are being compared should have similar noise values. The values detected here were between 1.59 and 3.06. For the *7.2sox10*:GFP samples most RawQ values were similar, ranging from 1.59 to 1.95 with the WT6 sample a slight outlier with a value of 2.65. For the *4.9sox10*:GFP arrays RawQ values ranged from 1.96 to 3.06 and were typically slightly higher than seen with the *7.2sox10*:GFP arrays (**Table 8** and **Table 9**).

The majority of quality control measures examined here displayed good characteristics. It was concluded that the sample preparation and hybridization protocols had been successful. This data gives confidence that meaningful results can be obtained from analyzing and comparing the array data that has been generated.

	Quality control Parameters	M1	M5	M6	WT2	WT4	WT6
Hybridization Controls (Present or Absent)	*bioB*	P	P	P	P	P	P
	bioC	P	P	P	P	P	P
	bioD	P	P	P	P	P	P
	cre	P	P	P	P	P	P
Poly-A controls (Present or Absent)	*lys*	P	P	P	P	P	P
	phe	P	P	P	P	P	P
	thr	P	P	P	P	P	P
	dap	P	P	P	P	P	P
3'/5' Ratio	β-actin	7.59	7.48	5.80	5.94	5.69	5.24
	GAPDH	1.48	1.20	1.21	1.47	1.55	1.21
	% Present	52.5	59.7	56.8	54.7	57.0	52.6
Background Signal	Average Background	73.94	48.63	54.08	55.64	49.98	76.39
Noise	RawQ	1.95	1.59	1.97	1.88	1.68	2.65

Table 8 : Summary of Quality control parameters for Arrays hybridized with *7.2sox10*:GFP derived samples.
Hybridization and Poly A controls are given a detection identifier based on signal values detected at 3' probes, P = Present. β-actin and GAPDH display a ratio of signal values of the 3' probe sets against the 5' probe sets. The percentage present shows the percentage of probes on the array detected as being present based on signal values. Average Background is a signal value (measure of intensity) for the average background detected across the whole array. RawQ is an Affymetrix measure of noise.

	Quality control Parameters	M3	M4	WT3	WT4
Hybridization Controls (Present or Absent)	*bioB*	P	P	P	P
	bioC	P	P	P	P
	bioD	P	P	P	P
	cre	P	P	P	P
Poly-A controls (Present or Absent)	*lys*	P	P	P	P
	phe	P	P	P	P
	thr	P	P	P	P
	dap	P	P	P	P
3'/5' Ratio	β-actin	6.76	8.77	9.57	8.84
	GAPDH	1.21	1.39	1.36	1.00
	% Present	59.5	54.0	45.4	52.0
Background	Average Background	55.4	89.45	58.29	74.85
Noise	RawQ	2.00	3.06	1.96	2.63

Table 9: Summary of Quality control parameters for Arrays hybridized with *4.9sox10*:GFP derived samples.
Presented as per Table 8.

3.2.5.2 Raw Array Data for Key Genes.

To assess whether differential gene expression could be detected from the array data, two samples (M6 and WT6) from the *7.2sox10*:GFP line were initially hybridized to the arrays. The results for several key genes from this WT versus mutant comparison were analysed (**Table 10**) prior to the remaining *7.2sox10*:GFP samples being hybridized to arrays. GCOS was used to generate a change call based on programmed statistical tests to compare these two arrays. *Sox10* and *foxd3* transcripts registered as unchanged when comparing these two arrays, for *sox10* this was expected but a decrease in *foxd3* expression was expected. The expected result may not have been detected because the sample size being examined was very small. The remaining key genes all displayed decreased expression in mutant samples as predicted. *Mitfa* registered as not expressed at all (Absent) in the mutant sample (**Table 10**). As the expected results for differential gene expression were detected the remaining samples were processed and hybridized to microarrays.

Affymetrix Identifier	Gene Name	WT6 signal	M6 signal	Change
Dr.12604.1.S1_at	*sox10*	3967.4 (P)	2940.3 (P)	NC
Dr.8080.1.S1_at	*mitfa*	115.4 (P)	34.5 (A)	D
Dr.10336.1.S1_at	*dct*	1454 (P)	386 (P)	D
Dr.19416.1.S1_at	*otomp*	326.7 (P)	154 (P)	D
Dr.590.1.S1_at	*foxd3*	198.2 (P)	165.9 (P)	NC
Dr.8124.1.S1_at	*crestin*	9940.8 (P)	4383.2 (P)	D
Dr.14668.1.S1_at	*gch*	1093.1 (P)	653.5 (P)	D
Dr.14668.2.S1_at	*gch*	2859.6 (P)	1456.6 (P)	D
Dr.9225.1.A1_at	*aox3* (*xdh*)	3824.9 (P)	3028 (P)	D

Table 10: Comparison of the raw microarray data for samples WT6 versus M6. Affymetrix identifiers were annotated using the NetAffx tool on the Affymetrix website to identify correct gene names. Raw data (signal values and detection calls) were then found for these genes from data analysed with GCOS. Comparison analysis was performed using GCOS following guidelines in the Affymetrix Data Analysis Fundamentals Handbook using default settings to generate change calls. The WT sample was set as the baseline array and the mutant array as the experimental sample. Detection calls, P = Present, A = Absent. Change calls, NC = No Change, D = Down.

Once all the *7.2sox10*:GFP samples had been hybridized to the arrays and the raw data received from the UCI Microarray Facility, it was examined. To confirm that decreases in transcript level for key genes was likely, signal intensity values for mutant samples versus WT samples were assessed (**Table 11**). All mutant samples had consistently lower mean transcript levels detected for all key genes examined. Particularly strong examples included *crestin* and *dct*. Both *sox10* and *foxd3* had lower average signal intensities in mutant samples, in contrast to the previous two array analysis, but this analysis did not include a statistical test so it cannot indicate if differences were significant or not. This analysis provided a simple inspection of the raw data and suggested that key marker genes for neural crest and ear development were down-regulated in *sox10* mutants. This result was similar to that seen when comparing the samples WT6 and M6

The same raw data shown in **Table 11** was compiled for the arrays hybridized with the *4.9sox10*:GFP samples (**Table 12**). Again the majority of key genes showed a reduction in transcript levels in the mutant samples, *crestin* and *dct* were strong examples of this. Transcript levels for *sox10* were actually slightly higher in mutant samples for this data set. This was in contrast to the *7.2sox10*:GFP analysis (**Table 11**) suggesting that *sox10* was not significantly changed. This was in agreement with the two array comparison shown in **Table 10**. Once again, examining the raw data suggested that key neural crest and ear genes were detected as down-regulated in *sox10* mutant embryos. The microarray data obtained fits with predicted results and should be analysed further to identify transcripts not previously associated with *sox10* biology.

Gene Name	WT2	WT4	WT6	Mean ± sd	M1	M5	M6	Mean ± sd
sox10	3425.9	3327.3	3967.4	3573.5 ± 344.6	2001.6	1949.7	2940.3	2297.2 557.5
mitfa	161.8	158.3	115.4	145.2 ± 25.8	44.9	32.2	34.5	37.2 ± 6.8
dct	1374.6	2553.7	1454	1794.1 ± 659.0	128.3	112.1	386	208.8 ± 153.7
otomp	218.2	387.8	326.7	310.9 ± 86.0	106.8	142.7	154	134.5 ± 24.6
foxd3	238.8	192.6	198.2	209.9 ± 25.2	190	164.1	165.9	173.3 ± 14.5
crestin	9227.4	8703	9940.8	9290.4 ± 621.3	2720.3	3339.5	4383.2	3481 ± 840.4
gch	1118.9	479.5	1093.1	897.2 ± 361.9	493.1	476.7	653.5	541.1 ± 97.7
gch	2505.3	1399.8	2859.6	2254.9 ± 761.4	1083.8	1159	1456.6	1233.1 ± 197.1
aox3	3803.6	3331.1	3824.9	3653.2 ± 279.2	2093.6	2386.5	3028	2502.7 ± 477.9

Table 11: Signal Intensities for key genes from the *7.2sox10*:GFP arrays compared between WT and mutant samples.
Signal intensities were generated using GCOS for each array and then the mean calculated for WT samples and mutant samples for each gene. Mean results are given ± standard deviation.

Gene Name	WT3	WT4	Mean ± sd	M3	M4	Mean ± sd
sox10	1897.9	2009.3	1953.6 ± 78.8	2457.5	1939.3	2198.4 ± 366.4
mitfa	134.7	156.6	145.65 ± 15.5	17.5	46.3	31.9 ± 20.4
dct	2891.3	2417.7	2654.5 ± 334.9	49.1	251.5	150.3 ± 143.1
otomp	155	128.5	141.75 ± 18.7	23.8	63.1	43.45 ± 27.8
foxd3	216.9	280.6	248.75 ± 45.0	118.5	150.2	134.35 ± 22.4
crestin	5178.2	6507.1	5842.65 ± 939.7	1646.6	2253.3	1949.95 ± 429.0
gch	1141.4	1143.0	1142.2 ± 1.1	380.1	471.5	425.8 ± 64.6
gch	2900	3170.9	3035.45 ± 191.6	1089.4	1620.6	1355 ± 375.6
aox3	3206.5	3870.8	3538.65 ± 469.7	1814.9	2368.6	2091.75 ± 391.5

Table 12: Signal Intensities for key genes from the *4.9sox10*:GFP arrays compared between WT and mutant samples.
Signal intensities were generated using GCOS for each array and then the mean ± standard deviation calculated for WT samples and mutant samples for each gene.

3.2.6 Normalization of Microarray data using GEPAS

To perform both the normalization of all the array data and to identify differential gene expression, tools freely available on the Gene Expression Profile Analysis Suite (GEPAS) website (www.gepas.org) were used (Montaner et al., 2006).

When dealing with multiple arrays, variation between arrays as a result of non-biological factors such as differences in sample preparation or array processing need to be eliminated. Such variation can obscure differential gene expression resulting from the biological factor being tested. Despite the quality control array parameters all being within acceptable limits, the arrays all had varying levels of background signal intensity (**Figure 14**) and it is elements such as this that must be normalized across all arrays being compared. To normalize the data, the *expresso* module on the GEPAS website was used. This module is based on the *affy* Bioconductor package (Gautier et al., 2004, Gentleman et al., 2004). It allows multiple different combinations of tools for correcting background, normalizing data, correcting perfect match (PM) probe intensities against mis-match (MM) probe intensities and summarization methods to be combined easily. At the end of this process two box plots were generated, one displaying the processed probe intensities and one showing probe intensities prior to processing (**Figure 14**). These can be used to facilitate the selection of the best combination of normalization tools. Based on this criterion the default GEPAS settings were selected that process the data as per the Robust multi-array analysis (RMA) expression measure (Irizarry et al., 2003b, Irizarry et al., 2003a). The settings used were as follows; Background correction = rma, Normalization = quantiles, PM correct methods = pmonly and the Summarization method = median polish. The data set generated from the *7.2sox10*:GFP samples and the *4.9sox10*:GFP samples were normalized individually and then combined and normalized to give a data set of five WT arrays versus five mutant arrays. The pre-normalization box plots show that the PM probe intensities of each set of arrays were not centred on the same point and did not have identical distributions. Post-normalization, each set of arrays became centred on the same point and showed very similar distributions of PM probe intensities (**Figure 14**). In conclusion, arrays were calibrated to the same scale and therefore comparisons between mutant and WT arrays in a set can be made successfully.

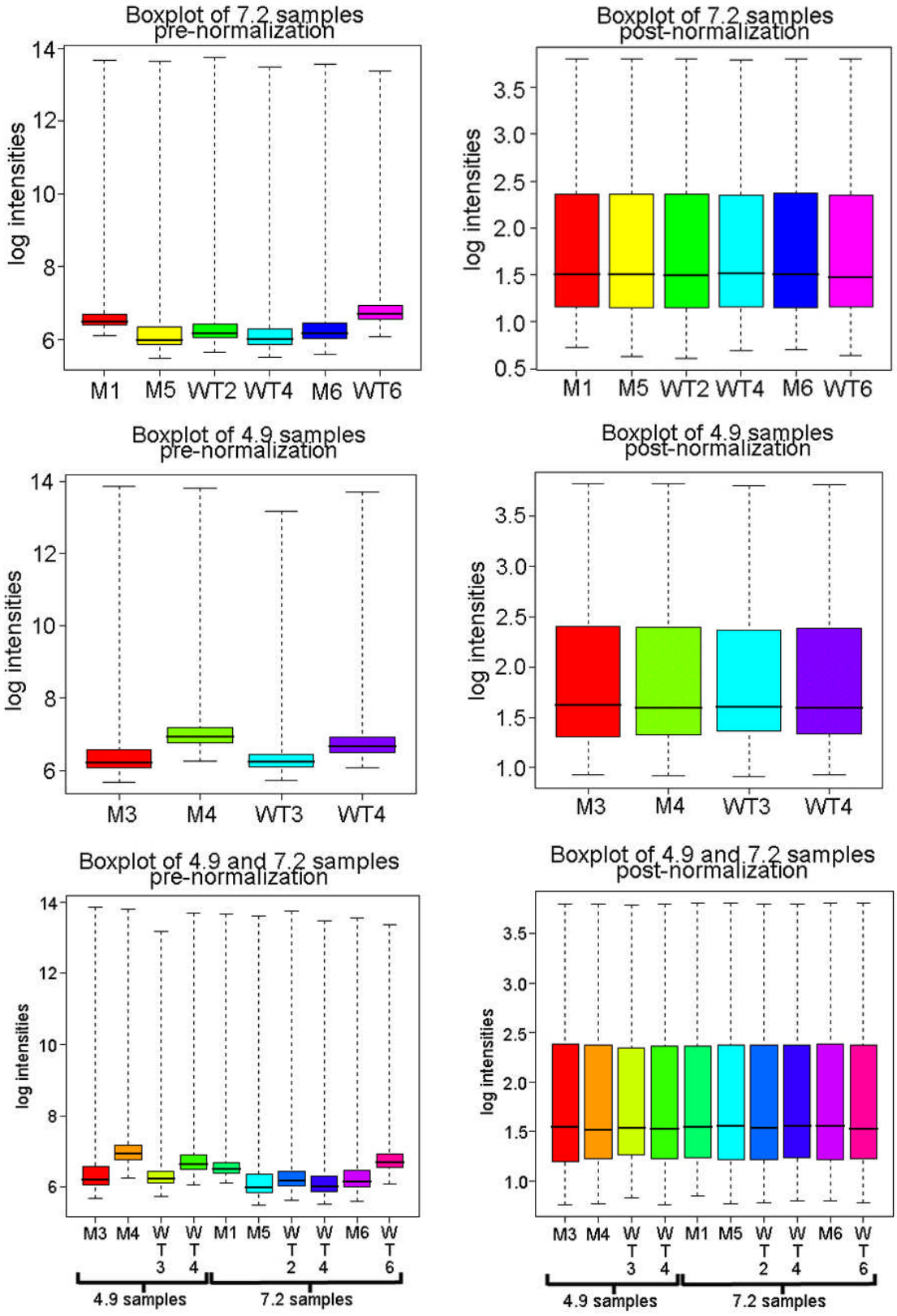


Figure 14: Box plots of probe intensities for raw and normalized array data.
Box plots on the left show the distribution of PM probe intensities for groups of array data, the top row shows *7.2sox10*:GFP data, the middle row shows *4.9sox10*:GFP data and the bottom row shows both these data sets combined as one. The box plots on the right side show the same PM probe intensity distribution after normalization by the RMA method. Probe intensities are shown in a $\log_2$ scale.

3.2.6 Analysis of Differential Gene Expression using GEPAS

To analyse the data normalized using *expresso*, another GEPAS tool was employed, *T-rex.* Output data from *expresso* can easily be entered into *T-rex*, but class identifiers needed to be added to distinguish mutant from WT data sets. The three normalized data sets (one set from each group of arrays and a set with both groups combined) were all analysed using the two class analysis package in *T-rex* primarily with the *t*-Test option. In addition the CLEAR test (Valls et al., 2008) option was also used with a significance level of 0.05. The *t*-Test generated a list of genes ranked by *t*-statistic with genes down-regulated in mutant embryos at the bottom of the list and genes up-regulated in mutant embryos at the top of the list. A *p*-value for the *t*-statistic was also reported. The *p*-value is a measure of the evidence against the null hypothesis, it is the probability of the test statistic being equal or greater than the observed result under the assumption that the null hypothesis is true (Cui and Churchill, 2003). If the *p*-value is smaller or equal to the defined significance level then the null hypothesis of no difference between the two classes can be rejected. The CLEAR test combines a *z*-test to identify large gene expression changes and a Chi squared test to evaluate variability (Valls et al., 2008). The CLEAR test was applied as it generated a value for mean difference between the two groups (m), when working with log intensities this is known as signal log ratio (D. Montaner personal communication), an output the *t*-test did not supply. A simple calculation can change the signal log ratio value into a fold change value (See Chapter 2). Additionally the CLEAR test in GEPAS identified the outcome of the test by text, for example whether a gene is differentially expressed (diff expr) or not changed (non. significant). This provides a clear and convenient readout of the test result.

All three sets of normalized data were analysed by *t*-Test and CLEAR test to identify genes differentially expressed in mutant embryos, both up-regulated and down-regulated genes. The total number of genes identified by each test is shown in **Table 13.** Using the *t*-test, large numbers of genes were identified as differentially expressed in all data sets at two different significance levels. GEPAS generated a graphical output of the most up-regulated and most down-regulated genes identified by the *t*-Test. Each gene was represented by a row and each array was represented by a column. The data for a gene on each array was represented by a coloured square to depict normalized expression data around a mean of zero. Red shows increased expression for a gene compared to the other sample group, blue shows decreased expression. As such a gene down-regulated in mutants will appear blue in the mutant sample and red in the WT sample. The more intense the colour the greater the expression difference based on the standardized scale displayed in sigma units in the output file. Each output file generated by the *T-rex* analyses is shown in **Figure 15** with key marker genes annotated. These grid images display the 50 most up-regulated genes in mutant samples at the top and the 50

most down-regulated genes at the bottom. The colours on the grid showed that the analyses have identified genes that are consistently differentially regulated in mutant samples across all arrays, red and blue colours were rarely mixed between mutant and WT classes for a gene. Mixing of colours was most common in the up-regulated gene list of the combined analysis dataset, showing that differential gene expression was not consistent across all arrays for some genes.

When correction methods for multiple testing were applied only a small number of genes in the combined dataset were significant, no significantly differentially regulated genes were identified in either single dataset at a significance level of $p \leq 0.05$ (**Table 13**). The Family Wise Error Rate (FWER) is a control on the probability of accumulating one or more false-positive (Type I) errors over a number of statistical tests. FWER criteria are very stringent to prevent false positives being identified and this can reduce the power of the test (probability that a real effect can be identified by a statistical test) (Cui and Churchill, 2003). In practice this will exclude a large number of true positives (Type II errors). The False Discovery Rate (FDR) defines the proportion of false positives among all of the genes identified as being differentially expressed (Cui and Churchill, 2003). At a defined FDR significance level a known proportion of genes will be false positives, this allows for a higher rate of false positives than the FWER correction and therefore can achieve more power than with FWER correction (Cui and Churchill, 2003). The *q*-value is similar to the *p*-value but is an extension of the FDR (Storey and Tibshirani, 2003), and hence it is not surprising that both FDR methods give identical results with the datasets.

The CLEAR test also identified a large number of differentially regulated genes in all data sets. The number of down-regulated genes identified in the separate 7.2 and 4.9 datasets by the CLEAR test was very similar in both cases to the number of genes identified by the *t*-Test at the $p \leq 0.05$ significance level. For the combined 7.2 + 4.9 dataset the CLEAR test identified 410 differentially regulated genes, over two hundred fewer genes than the *t*-Test at the $p \leq 0.05$ significance level. The CLEAR test also identified much larger numbers of up-regulated genes than the *t*-Test at the $p \leq 0.05$ significance level.

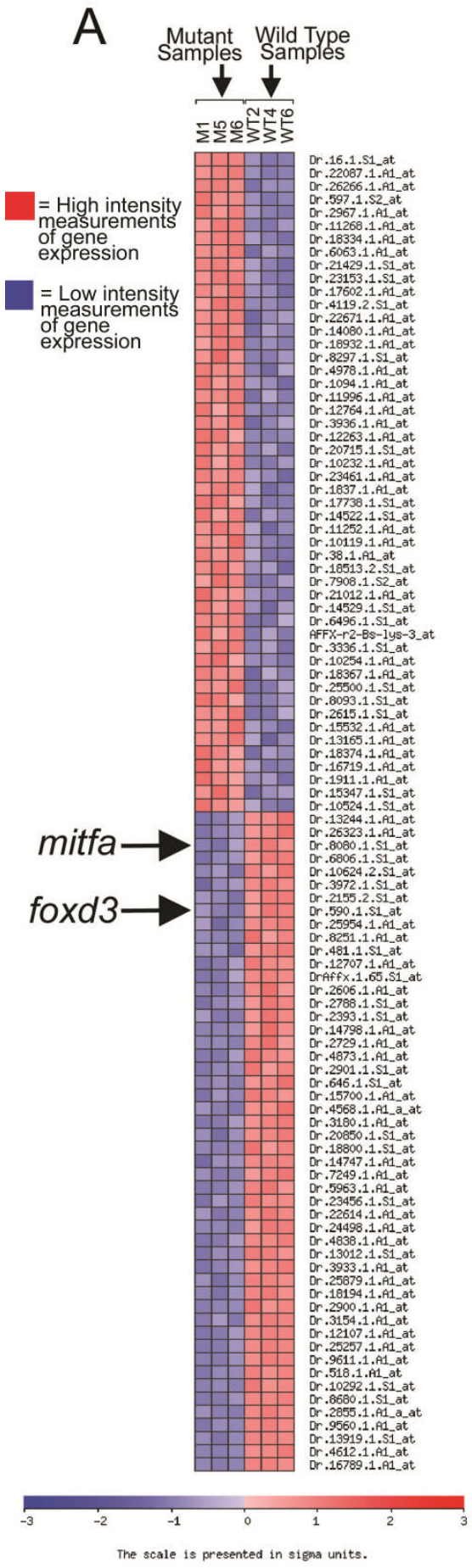
Having generated lists of differentially expressed genes, they were examined to see how the key marker genes performed (**Table 14**). Neither statistical test identified *sox10* as differentially regulated in *sox10* mutant embryos, the expected result. All other marker genes in every dataset were identified as down-regulated by the CLEAR test. The *t*-Test for the 7.2 samples generally performed well but neither *gch* probe was identified as being differentially regulated, perhaps due to the high variability identified by the CLEAR test. The *t*-Test produced mixed results for the 4.9 samples failing to declare *mitfa*, *foxd3* or *crestin* as being differentially regulated but successfully identifying both *gch* probe sets

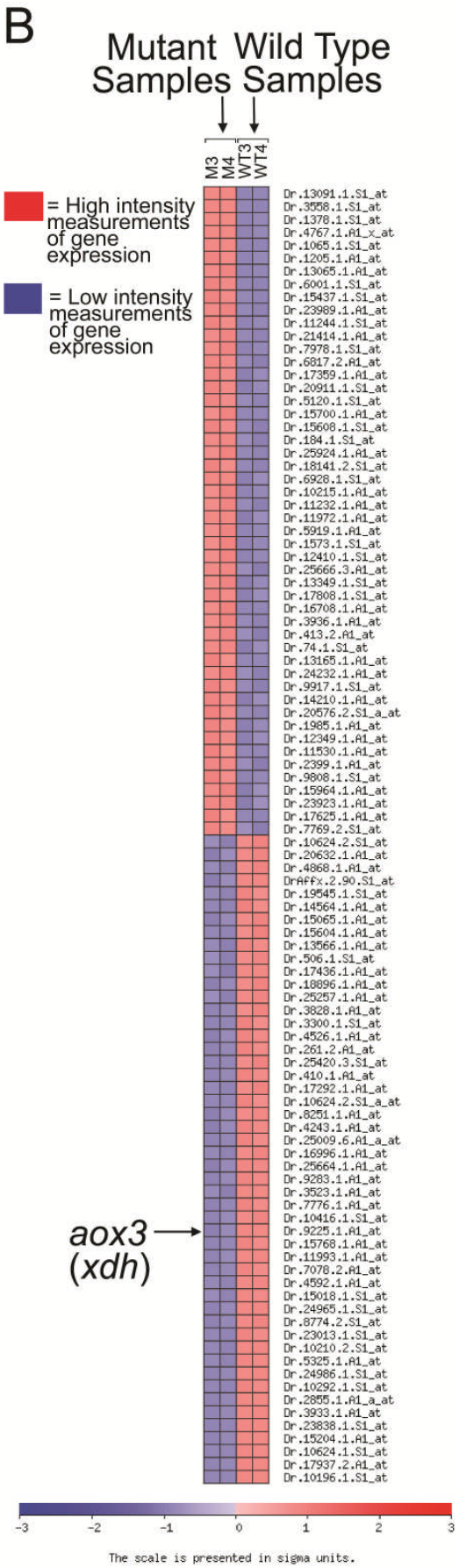
as being down-regulated. The *t*-Test on the combined 7.2 and 4.9 analysis performed well identifying all marker genes other than *sox10* as being down-regulated. Differential gene expression between mutant and WT samples was successfully identified by microarray analysis by both statistical tests.

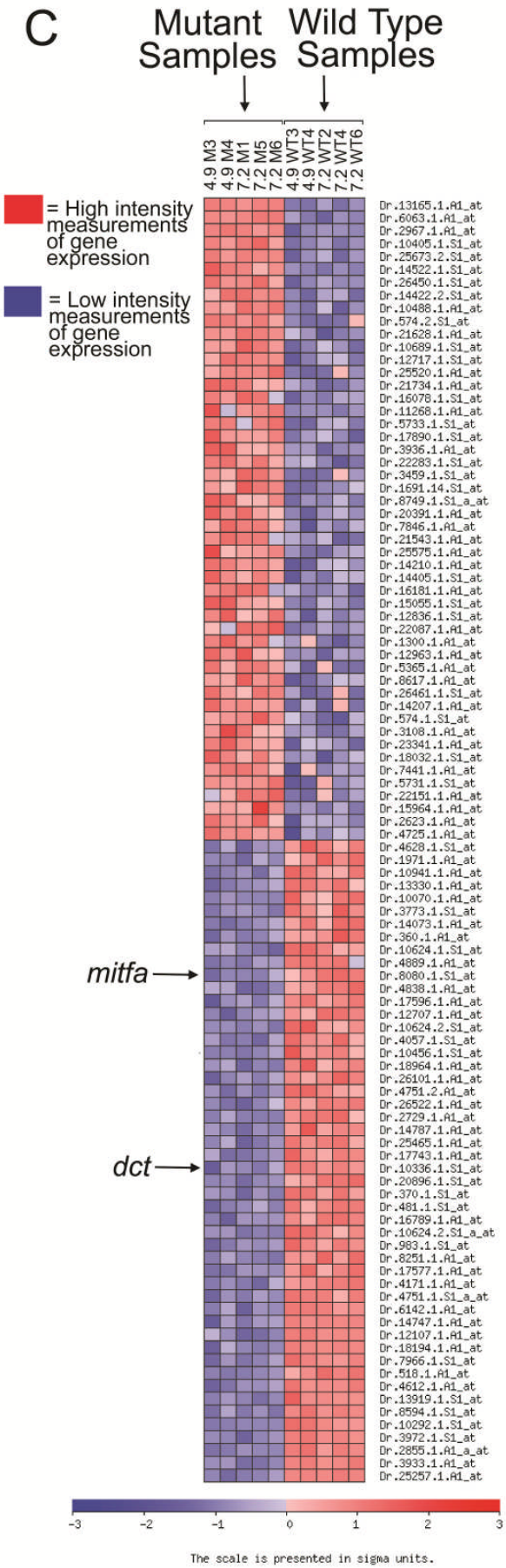
	t-Test					CLEAR Test		
Down	$p \leq 0.05$	$p \leq 0.01$	FWER ($p \leq 0.05$)	FDR ($p \leq 0.05$)	*q*-value ($q \leq 0.05$)	Diff Expr	Diff Expr High Var	Total
7.2	694	200	0	0	0	496	109	605
4.9	502	121	0	0	0	354	154	508
7.2+4.9	640	238	6	32	32	308	102	410
Up	$p \leq 0.05$	$p \leq 0.01$	FWER ($p \leq 0.05$)	FDR ($p \leq 0.05$)	*q*-value ($q \leq 0.05$)	Diff Expr	Diff Expr High Variability	Total
7.2	577	121	0	0	0	566	308	874
4.9	455	91	0	0	0	578	222	800
7.2+4.9	494	133	2	7	7	359	351	710

Table 13: Summary of the number of genes identified as differentially expressed.
For all three datasets the number of differentially regulated genes, either up or down, identified have been summarised from the *t*-Test and the CLEAR test. The *t*-Test shows the results for two different significant levels and the FWER, FDR and *q*-value corrections for multiple testing that were applied. The CLEAR test shows genes identified as differentially expressed (Diff Expr) and differentially expressed but with a high variability (Diff Expr High Var) and the total for these when combined.

Figure 15: Grid image output files representing *T-rex* differential gene expression *t*-Test analyses.
All grid images are organised with arrays in columns and genes in rows, mutant sample arrays are on the left and WT sample arrays are on the right. Red = high intensity gene expression, blue = low intensity gene expression with darker shades indicating increasing extremes of high or low expression when compared to the other class. The 50 most up-regulated genes are shown in the top half of the grid, the 50 most down-regulated genes in the bottom half of the grid. (A) Shows the 7.2 sample array data, (B) shows the 4.9 sample array data and (C) shows the combined 7.2 + 4.9 samples array data. The data is shown on a standardized scale in sigma units. Key marker genes are indicated. (See next page)


A
Mutant Samples
Wild Type Samples
M1
M5
M6
WT2
WT4
WT6
= High intensity measurements of gene expression
= Low intensity measurements of gene expression
mitfa
foxd3
-3
-2
-1
0
1
2
3
The scale is presented in sigma units.


B
Mutant Samples
Wild Type Samples
M3
M4
WT3
WT4
= High intensity measurements of gene expression
= Low intensity measurements of gene expression
aox3 (xdh)
Dr.13091.1.S1_at
Dr.3558.1.S1_at
Dr.1378.1.S1_at
Dr.4767.1.A1_x_at
Dr.1065.1.S1_at
Dr.1205.1.A1_at
Dr.13065.1.A1_at
Dr.6001.1.S1_at
Dr.15437.1.S1_at
Dr.23989.1.A1_at
Dr.11244.1.S1_at
Dr.21414.1.A1_at
Dr.7978.1.S1_at
Dr.6817.2.A1_at
Dr.17359.1.A1_at
Dr.20911.1.S1_at
Dr.5120.1.S1_at
Dr.15700.1.A1_at
Dr.15608.1.S1_at
Dr.184.1.S1_at
Dr.25924.1.A1_at
Dr.18141.2.S1_at
Dr.6928.1.S1_at
Dr.10215.1.A1_at
Dr.11232.1.A1_at
Dr.11972.1.A1_at
Dr.5919.1.A1_at
Dr.1573.1.S1_at
Dr.12410.1.S1_at
Dr.25666.3.A1_at
Dr.13349.1.S1_at
Dr.17808.1.S1_at
Dr.16708.1.A1_at
Dr.3936.1.A1_at
Dr.413.2.A1_at
Dr.74.1.S1_at
Dr.13165.1.A1_at
Dr.24232.1.A1_at
Dr.9917.1.S1_at
Dr.14210.1.A1_at
Dr.20576.2.S1_a_at
Dr.1985.1.A1_at
Dr.12349.1.A1_at
Dr.11530.1.A1_at
Dr.2399.1.A1_at
Dr.9808.1.S1_at
Dr.15964.1.A1_at
Dr.23923.1.A1_at
Dr.17625.1.A1_at
Dr.7769.2.S1_at
Dr.10624.2.S1_at
Dr.20632.1.A1_at
Dr.4868.1.A1_at
DrAffx.2.90.S1_at
Dr.19545.1.S1_at
Dr.14564.1.A1_at
Dr.15065.1.A1_at
Dr.15604.1.A1_at
Dr.13566.1.A1_at
Dr.506.1.S1_at
Dr.17436.1.A1_at
Dr.18896.1.A1_at
Dr.25257.1.A1_at
Dr.3828.1.A1_at
Dr.3300.1.S1_at
Dr.4526.1.A1_at
Dr.261.2.A1_at
Dr.25420.3.S1_at
Dr.410.1.A1_at
Dr.17292.1.A1_at
Dr.10624.2.S1_a_at
Dr.8251.1.A1_at
Dr.4243.1.A1_at
Dr.25009.6.A1_a_at
Dr.16996.1.A1_at
Dr.25664.1.A1_at
Dr.9283.1.A1_at
Dr.3523.1.A1_at
Dr.7776.1.A1_at
Dr.10416.1.S1_at
Dr.9225.1.A1_at
Dr.15768.1.A1_at
Dr.11993.1.A1_at
Dr.7078.2.A1_at
Dr.4592.1.A1_at
Dr.15018.1.S1_at
Dr.24965.1.S1_at
Dr.8774.2.S1_at
Dr.23013.1.S1_at
Dr.10210.2.S1_at
Dr.5325.1.A1_at
Dr.24986.1.S1_at
Dr.10292.1.S1_at
Dr.2855.1.A1_a_at
Dr.3933.1.A1_at
Dr.23838.1.S1_at
Dr.15204.1.A1_at
Dr.10624.1.S1_at
Dr.17937.2.A1_at
Dr.10196.1.S1_at
-3
-2
-1
0
1
2
3
The scale is presented in sigma units.


C
Mutant Samples
Wild Type Samples
4.9 M3
4.9 M4
7.2 M1
7.2 M5
7.2 M6
4.9 WT3
4.9 WT4
7.2 WT2
7.2 WT4
7.2 WT6
= High intensity measurements of gene expression
= Low intensity measurements of gene expression
mitfa
dct
-3
-2
-1
0
1
2
3
The scale is presented in sigma units.

Gene	Data Set	*p*-value	CLEAR result	Fold Change
sox10 Dr.12604.1.S1_at	7.2	0.1320938021	non.significant	-1.16
	4.9	0.5171018839	non.significant	1.07
	7.2 + 4.9	0.3415146768	non.significant	-1.07
mitfa Dr.8080.1.S1_at	7.2	0.0013052389	diff.expr	-2.14
	4.9	0.0613204651	diff.expr	-1.65
	7.2 + 4.9	0.0001868201	diff.expr	-1.89
dct Dr.10336.1.S1_at	7.2	0.0022070811	diff.expr.high.var	-9.59
	4.9	0.0318133533	diff.expr.high.var	-26.94
	7.2 + 4.9	0.0000556806	diff.expr.high.var	-14.58
otomp Dr.19416.1.S1_at	7.2	0.0026735307	diff.expr	-1.86
	4.9	0.0119594019	diff.expr	-1.49
	7.2 + 4.9	0.0160892177	diff.expr.high.var	-1.96
foxd3 Dr.590.1.S1_at	7.2	0.0010371582	diff.expr	-1.32
	4.9	0.1877090707	diff.expr	-1.50
	7.2 + 4.9	0.0126530966	diff.expr	-1.32
crestin Dr.8124.1.S1_at	7.2	0.0015099111	diff.expr	-2.49
	4.9	0.0867269561	diff.expr	-2.57
	7.2 + 4.9	0.0018713886	diff.expr.high.var	-2.53
gch Dr.14668.1.S1_at	7.2	0.0916629434	diff.expr.high.var	-1.87
	4.9	0.0202664994	diff.expr	-2.41
	7.2 + 4.9	0.0032915222	diff.expr.high.var	-2.11
gch Dr.14668.2.S1_at	7.2	0.1258791536	diff.expr.high.var	-1.69
	4.9	0.0314634927	diff.expr	-2.54
	7.2 + 4.9	0.0044618039	diff.expr	-1.97
aox3 (*xdh*) Dr.9225.1.A1_at	7.2	0.0139850015	diff.expr	-1.52
	4.9	0.0016299261	diff.expr	-1.59
	7.2 + 4.9	0.0010224523	diff.expr	-1.50

Table 14: Summary of T-rex differential gene expression analysis for key marker genes.

Both the *t*-Test *p*-value and the CLEAR test results for all three datasets are shown. The fold change readout generated from the CLEAR test m value is also presented. Results identifying a gene as differentially regulated are highlighted in green, results identifying a gene as not differentially regulated are highlighted in red. For the *t*-Test a *p*-value significance level of $p \leq 0.05$ was selected. CLEAR test results indicate the following outcomes: Non significant (not differentially expressed), diff expr (differentially expressed) and diff expr high variability (differentially expressed with high variability).

3.3 Discussion

3.3.1 On FACS of zebrafish GFP positive cells

The use of FACS to enrich for GFP positive cells from the transgenic embryos resulted in selecting approximately between 1.5 % and 3 % (**Table 5** and **Table 6**) of the total number of disaggregated cells. As expected, a very small number of cells were purified from the total number derived from disaggregated embryos. Importantly, key NC and ear marker genes were enriched in GFP positive FACS samples when compared to GFP negative samples by RT-PCR (**Figure 9**). Thus FACS in conjunction with the *7.2sox10*:GFP and *4.9sox10*:GFP transgenic lines proved to be capable of enriching for the cell types of interest.

Whereas NCCs were expected to be enriched from both transgenic lines, otic epithelium cells were only expected to be enriched from the *7.2sox10*:GFP line. Thus the percentage of GFP positive cells purified from WT embryos of the *4.9sox10*:GFP line should provide the most accurate measure of the percentage of NCCs in an embryo. The results show this to be approximately 1.5 % to 1.75 % cells in an embryo (**Table 6**). This is likely to be an overestimate as some otic epithelium cells and muscle cells expressing GFP (**Table 4**) would also be purified. In fact *otomp* was enriched in GFP positive samples from both transgenic lines when compared to GFP negative samples (**Figure 9**). Therefore GFP positive otic epithelium cells were sorted from the *4.9sox10*:GFP line. It was expected that due to reduced fluorescence in the otic epithelium of the *4.9sox10*:GFP line, considerably fewer of these cells would be selected. In line with this, a *t*-Test to compare the number of cells purified from each transgenic line demonstrated that there were significantly fewer cells purified per 500,000 cell FACS run from the *4.9sox10*:GFP line. In addition *otomp* appeared to be less enriched from GFP positive *4.9sox10*:GFP cells than from *7.2sox10*:GFP cells as shown by RT-PCR (**Figure 9**). Comparison of *otomp* expression between WT and mutant *4.9sox10*:GFP samples still identified the expected down-regulation in *sox10* mutant samples (**Figure 11**). This demonstrates that otic epithelium cells have been purified from the *4.9sox10*:GFP line and that ear marker genes can still be identified from this data set. It will be interesting to observe how any other otic epithelium expressed genes behave when compared between *7.2sox10*:GFP and *4.9sox10*:GFP data sets. This data does suggest that it will be difficult to select for *sox10* regulated ear specific genes by comparing microarray results generated from the two transgenic lines.

It was noticed when purifying GFP positive cells from mutant samples that more cells were selected per 500,000 cells than from WT samples. For the *7.2sox10*:GFP line this difference was statistically significant as determined by a *t*-Test. For the *4.9sox10*:GFP line this proposed difference was very close to being significant with the *t*-

Test generating a *p*-value of 0.054. These results hint that there could be proportionately more GFP positive cells in mutant embryos than in WT embryos. This could have occurred as a result of passing WT samples through the FACS machine after mutant samples, the increased time spent on ice after embryonic disaggregation may have affected GFP positive cell number. There may have been some cell death occurring with time after disaggregation, but cell death was not quantified in either WT or mutant samples or over a time course. Additionally the expectation would be for cell death to affect GFP positive and GFP negative cells equally. Alternatively some *sox10*$^{-/-}$ cells could be maintaining GFP expression as a result of failing to specify correctly. It has been noted that *sox10* is required for neural crest multipotentiality (Kim et al., 2003) and that down-regulation of *sox10* occurs during melanophore differentiation (Greenhill, 2008) and sensory neuron differentiation (Carney et al., 2006). Thus in *sox10* mutants, where differentiation does not occur due to NCCs failing to become specified (Dutton et al., 2001a) *sox10* expression may persist. This could result in a number of additional GFP positive cells being present in mutant embryos compared to WT embryos. However *sox10* expression is being visualised indirectly by the presence of GFP protein which can persist beyond active transcription thus this may not explain the possible difference in GFP positive cell number between WT and *sox10* mutant embryos.

Enrichment of fluorescent cells by FACS presents a powerful tool to generate a pure population of cells for gene expression analysis. However if the process of embryo disaggregation and passage through a FACS machine alters the gene expression profile of cells then aberrant results may be obtained. Examination of signal intensities from raw microarray data (**Table 11** and **Table 12**) showed that key NC marker genes were all detected as present in WT samples suggesting that the expression profile of NCCs was intact. Additionally, down-regulation of key marker genes between WT and *sox10* mutant samples (**Table 10** and **Table 14**) was observed as expected thus further suggesting that cells retained their endogenous gene expression profile. Having only examined one ear specific gene, *otomp*, there is not enough information to gauge if the gene expression profile of otic epithelium cells has been affected. However *otomp* was detected as expressed in the sorted samples and appeared to be down-regulated in mutant samples thus behaving as expected. If the gene expression profile of NCCs is unaffected by FACS then that of otic epithelium cells would be expected to be normal as well.

Thus far all results suggested that GFP positive cells of interest have been successfully enriched by FACS and that the purified cells have normal gene expression profiles. This included expected differences in gene expression when compared between WT and mutant cells. However refinements to an experiment are always possible. Only two parameters, size (FSC) and fluorescence were used characterise the cell population of interest during cell sorting. While GFP expression faithfully marks the cell types of

interest, additional parameters could have been examined. Side scatter (SSC) provides a measure of cell complexity (granularity) and could have been employed to further define the cellular population of interest. However the granularity of our cell types of interest has not been examined and could therefore be misleading. Parameters such as FSC and fluorescence can be combined with a measure of pulse width which helps to define single cells (Cerda et al., 2009). Thus a population of fluorescent single cells could be enriched. A wide range of GFP positive events were sorted based on FSC (**Figure 8**). It is possible that clumps of cells could have been sorted as GFP positive based on the fluorescence of one cell thus contaminating the sample with GFP negative cells. Additionally a stain such as propidium iodide could be added to a sample. Propidium iodide is a fluorescent dye that binds DNA however it is excluded from live cells thus it marks dead cells for elimination from FACS (Cerda et al., 2009, Zhu et al., 2005). Cerda et al., 2009 demonstrated that only 1 % of their single cell population was constituted of dead cells. This is a very small percentage and thus FACS seems to be a reasonably gentle procedure. The number of dead cells in the samples presented here was not quantified to confirm it is true for this protocol. It is also possible to run a sample of cells from non-transgenic embryos through a FACS machine to evaluate background fluorescence and thus clearly define GFP negative cells for exclusion (Reeves and Posakony, 2005, Cerda et al., 2009). Having sorted a population of interest it is possible to re-run a sample of sorted fluorescent cells through a FACS machine to demonstrate the efficacy of the sorting process (Covassin et al., 2006). This can identify the number of GFP negative cells incorrectly sorted during the initial run. These additional measures help to build confidence that the cell sorting protocol being followed is effective, certainly it would be sensible to incorporate some of these controls in future experiments. The efficacy of the sorting protocol can at least partly be judged by the results of RT-PCR of sorted GFP positive versus GFP negative RNA samples (**Figure 9**). For the key marker genes assessed so far results indicate that the sorting protocol has been effective. To ensure the purity of sorted material a second sort can be performed (Zou et al., 2006) however in this case the cells collected were not processed for gene expression analysis. It would need to be demonstrated that cells sorted twice were not detrimentally affected by this leading to an altered gene expression profile. The protocol presented in this report was designed so that cells were exposed to minimal manipulation and that this manipulation was performed as quickly as possible to ensure that the gene expression profile of the cells of interest was maintained. Examination of gene expression by both RT-PCR (**Figure 9**, **Figure 10** and **Figure 11**) and during initial microarray analysis steps (**Table 11**, **Table 12** and **Table 14**) all indicate to some extent that this has been achieved. A considerable number of real targets of *sox10* were identified from the microarray screen by *in situ* hybridization (see Chapter 4) thus indicating the effectiveness of the experiments performed to prepare samples for microarray analysis.

3.3.2 Regarding the quantification and validation of extracted RNA

RNA was extracted from the sorted cell samples using the Qiagen RNeasy small sample extraction kit but due to the low concentration and small amount of RNA extracted in each sample they could not be accurately quantified using a spectrophotometer. Samples had to be sent to collaborators at the UCI DNA protein and microarray facility to accurately assess the quantity and quality of the extracted RNA using an Agilent Bioanalyser. As a result of this, some degraded RNA samples such as 4.9_M1 and 4.9 _M2 (**Figure 13**) were sent to the USA that were unsuitable for microarray analysis. RT-PCR analysis of the extracted RNA samples gave no indication of degradation, both 4.9_M1 and 4.9_M2 samples show successful amplification of both *sox10* and *actin* transcripts. Of course it is possible that degradation occurred during transport on dry ice to the USA after RT-PCR analysis was performed. It should also be noted that the *actin* primers used are positioned around an intron and therefore should amplify a larger band from genomic DNA than from cDNA. No genomic contamination was detected in any of the samples however several samples, especially 7.2_M3 and 4.9_WT1, had high levels of genomic contamination as determined by the Bioanalyser. This inability to accurately check the RNA samples hindered the progression of the project. Indeed samples from 48 hpf embryos were sorted from both WT and mutant embryos by FACS and RNA was extracted for microarray analysis. These samples were sent to the USA only to discover that the samples were badly degraded. While RT-PCR has proved suitable to give an indication of differential gene expression, it appears to be an unsuitable guide to other quality issues. Certainly access to an Agilent Bioanalyser for accurate RNA assessment would have been beneficial.

RT-PCR analysis of the extracted RNA gave an indication that RNA from WT and mutant samples displayed differential gene expression for two key NC genes, *mitfa* and *dct*. However this difference could only be judged semi-quantitatively, a quantitative approach such as Quantitative RT-PCR (QRT-PCR) would have defined the difference in transcript levels between the samples more accurately. A QRT-PCR approach was not pursued as this technique is not established in the lab and a simple RT-PCR approach enabled the project to proceed more rapidly.

3.3.3 Examining the Affymetrix Quality Control and raw data.

Guidelines for assessing Affymetrix GeneChips were followed as per the Affymetrix Data Analysis Fundamentals Handbook (Affymetrix, 2008) and results are shown in (**Table 8** and **Table 9**). The poly-A RNA and hybridization controls were all detected as present, the expected result. The *trp* poly-A control was not spiked into the samples by the UCI DNA and Protein Microarray facility and therefore was not detected (data not shown). The

average background results were all within the normal range defined by Affymetrix and thus were acceptable. The noise or RawQ values detected varied from array to array however these were normalised subsequent to this analysis using the GEPAS *expresso* tool. The percentage present analysis consistently showed 50 – 60 % of probes on each array to be called present, this seems to be relatively consistent for sets of data examining gene expression for the same population of GFP positive cells regardless of mutant or WT genetic status (**Table 8** and **Table 9**). The majority of genes that these cells express, such as key metabolic and housekeeping genes, should be the same and detected as present. The 4.9_WT3 array only showed 45 % of probes as present, while this appears to be a slight outlier it is not vastly different from the other arrays. The use of percent present as a measure for assessing array quality control should be used in conjunction with other parameters (Affymetrix, 2008). With other parameters performing well for the 4.9_WT3 array this slightly lower percent present value does not indicate a problem. For all arrays a high 3'/5' value was detected for the actin probes while the GAPDH probes showed ratios below the maximum of three specified by Affymetrix. A high 3'/5' ratio can indicate that that RNA sample was degraded or that the *in vitro* transcription reaction steps were inefficient. The 3'/5' ratio can be increased by rounds of amplification (Affymetrix, 2008) and the samples have undergone linear amplification, but this might be expected to affect both internal control genes. The high actin 3'/5' ratio may result from use of the NuGen Ovation™ Biotin Pico RNA Amplification and Labelling system kits to perform the RNA amplification and labelling steps rather than Affymetrix supplied kits. Indeed the 3'/5' ratios for actin and GAPDH observed here are very similar to other studies using NuGen kits including the actin ratios being high (Clément-Ziza et al., 2009). Additionally some experimental parameters have been altered at the UCI DNA and Protein Microarray Facility. A combination of these changes may cause different numbers for some quality control measures to be registered (Seung-Ah Chung, personal communication). Despite some abnormal quality control measures being detected, the majority were acceptable thus giving confidence that the array sample processing and hybridization steps have performed well.

The true test of array performance comes from deriving biologically relevant results from the data. The performances of several key NC genes and an otic epithelium marker gene were assessed to validate the raw array data (**Table 10**, **Table 11** and **Table 12**). Raw signal values for all marker genes displayed a reduction in mutant samples thus recapitulating results seen from *in situ* hybridization patterns (**Figure 6**). As expected *sox10* expression levels remained unchanged. The majority of these key marker genes were also significantly down-regulated when the 7.2_M6 and 7.2_WT6 arrays were compared using GCOS, the only exception being *foxd3*. These results indicate that

biologically significant results can be derived from the data and therefore further analysis of the data to identify new *sox10* targets will be valuable.

3.3.4 Regarding the normalization and analysis of microarray data using GEPAS.

Based on trials using various combinations of options available in the GEPAS tool *expresso* and reading of the literature, the RMA expression measure was selected for processing the data sets. The RMA expression measure has been compared favourably to other normalization methods, such as those employed by Affymetrix, including a greater sensitivity in detection of differential gene expression (Irizarry et al., 2003a) and a higher correlation with real time RT-PCR data used to validate differential gene expression (Millenaar et al., 2006). The RMA expression measure uses a global model of probe intensity distribution based on all the array data sets to correct PM probes only (Irizarry et al., 2003b). The background correction calibrates the scales of all the arrays to bring them all to the same base level. As such this tries to eliminate non-biological factors, for example differences in the surface of the array that affect probe intensities on each individual array (www.gepas.org). Subsequently quantile normalization was performed in order to convert the distribution of probe intensities for each array in a set to the same scale (Bolstad et al., 2003). This step removes non-biological variation across a group of arrays by calibrating them to the same scale. Quantile normalization attempts to remove variation resulting from different scanner settings or laboratory conditions (www.gepas.org). As only PM probes are adjusted by the RMA method, no PM correction method based on MM probe intensity was applied. Millenaar *et. al.* 2006 showed that MM probe intensities often correlate with PM probe intensities and thus do not solely represent non-specific signal and their use can therefore lead to an underestimate of the true expression signal. Finally medianpolish summarization generates expression values in a $\log_2$ scale (Irizarry et al., 2003b). Summarization combines all the probe intensities that interrogate a single transcript (a probe set), into a single value to reflect the expression of that transcript (www.gepas.org). It is appropriate to log transform expression values as "hybridized probe intensities tend to be distributed over exponential space due to hybridization behaviour that is governed by exponential functions of sequence-dependent base pairing energetic" (Affymetrix, 2008). If signal values are plotted on a linear scale, high or intense data points mask the variation in less intense low values. Log transforming this data enables variations in both high and low value data points to be seen (Affymetrix, 2008). Despite the advantages discussed above it would be interesting to compare different normalization methods available on GEPAS using key marker genes to discover the process which most successfully identifies differentially expressed genes when subsequently analysed using the *T-rex* tool. The additional marker genes identified during the validation screen in Chapter 4 would make such analyses more comprehensive.

Having normalized the microarray data sets with the RMA method in *expresso*, the *T-rex* tool was utilised to identify differential gene expression using both the *t*-Test and the CLEAR test statistical test options. Both performed well when checking the results using the genes known to be down-regulated in *sox10*$^{-/-}$ embryos. The *sox10* mutant *m618* allele used for this screen expresses a full length *sox10* transcript but a single amino acid change renders the protein effectively non-functional (Dutton et al., 2001a). However, as a transcript that will hybridize to a microarray is still produced; no difference in *sox10* expression was expected. In fact, the CLEAR test outperformed the *t*-Test (**Table 14**). However it was the *t*-Test statistic that was used to generate the list of potential *sox10* targets for validation. The *t*-Test was preferred as it is in common use and microarray results using the *t*-Test have been published (Tanaka et al., 2000, Callow et al., 2000, Covassin et al., 2006). The *t*-Test can identify genes with very small differences in gene expression levels; however the test can suffer from low power if each gene only has a few replicates (Baldi and Long, 2001, Cui and Churchill, 2003). When only a few tests are performed, a gene with a very small difference in gene expression may appear to be a very significant hit if variance is also low thus generating false positives (Valls et al., 2008, Cui and Churchill, 2003). The *t*-Test performed best when applied to the combined data set where each gene is represented by 5 arrays, by increasing the number of replicates the power of the test increases. Another common method to identify differentially expressed genes is to evaluate fold change. This technique does not generate any statistical test to indicate a level of confidence in the classification of a gene as differentially regulated and can be biased towards identifying genes with low intensities as differentially expressed (Cui and Churchill, 2003). No information on the CLEAR test had been published at the time the data analysis was performed (Valls et al., 2008). Consequently not enough information was available to select the test as the primary tool to identify differential gene expression. A recent paper suggests the CLEAR test has more power to detect large changes in differential expression, but can fail to identify genes that display small changes in expression (Valls et al., 2008). This may disadvantage the CLEAR test in terms of identifying transcription factor genes which often have low expression levels. Nevertheless the CLEAR test did successfully identify *mitfa* and *foxd3*, two transcription factors, as differentially expressed (**Table 14**). There are additional statistical tests available in the *T-rex* tool to identify differential gene expression that could be applied to the data (Montaner et al., 2006, Tárraga et al., 2008). The success of the analysis will have to be judged on the number of targets that are successfully validated. Once an increased set of known differentially regulated genes is established it will provide an excellent tool to reassess the best route for data normalization and differential gene expression analysis. Identifying the best pipeline for these processes may aid in generating a list of highly probable *sox10* targets for data mining.

The GEPAS *t*-Test generates a *t*-statistic and then computes a *p*-value; genes falling below a defined *p*-value level are regarded as significant. When performing microarray data analysis thousands of tests (one for each gene on the array) are conducted leading to an accumulation of false positives. To address this, methods to correct for multiple testing can be applied. When applying either the FWER, FDR or *q*-value correction methods the majority of the genes were no longer considered to be significantly differentially expressed. This includes genes demonstrated by *in situ* hybridization to be down-regulated in *sox10* mutant embryos. The genes identified as not differentially expressed when multiple testing correction methods were applied contains biologically relevant results. As such the multiple testing correction methods have been ignored for the purposes of the microarray experiments. This increases the probability that in the list of genes significant by *p*-value alone, a number of false positive results are present. Additionally it is interesting to note that the only genes found to be significant after corrections for multiple testing are applied are within the combined dataset, the data set with the largest number of replicate experiments per gene. This suggests that comparing three WT versus three mutant arrays is not a large enough sample size to eliminate or at least to reduce the probability of false positive errors. Again this relates to the performance of the *t*-Test which improves with increasing replicate experiments for each gene. The genes considered to be significantly differentially expressed after applying a multiple testing correction method would be worth examining as it would be predicted that this list of genes contains very few false positives (see Chapter 5). This analysis highlights the issue of the sample size required in microarray experiments to generate statistically significant results. This is likely to be dependent on the biological sample being studied. However this project has also shown that some biologically significant results can be identified by only comparing 2 WT and 2 mutant arrays (**Table 14**) but by increasing sample size results more consistent with the true biological situation are generated (**Table 14**). A larger sample size is likely to reduce false positive results and improve the quality of generated data; this in turn will result in easier downstream analysis and more true positives being validated.

3.3.5 Alternative Strategies for Identifying *Sox10* targets.

The microarray approach adopted in this project was designed to identify any gene differentially expressed in *sox10* mutant embryos in comparison to WT embryos, including direct and indirect targets of Sox10. An alternative strategy that could be employed to identify *sox10* targets is a chromatin immunoprecipitation (ChIP) assay. This approach could be applied to whole embryos or GFP positive cells purified by FACS from transgenic lines. This technique uses an antibody specific to the protein of interest (in this case Sox10) to precipitate out the protein while bound to its native target DNA sequences.

These DNA sequences can then be hybridized to a microarray of genomic DNA or promoter sequence probes enabling the identification of the promoter sequences enriched by the precipitation step; together this approach is known as ChIP-on-chip. ChIP has been used as part of a pipeline to validate direct targets of Sox10 that were identified through addition of a small interfering RNA (siRNA) specific for Sox10 to a Schwannoma cell line (Lee et al., 2008). Microarray analysis was used to identify down-regulated genes and four genes were then validated as novel direct targets using ChIP (Lee et al., 2008). To successfully apply this technique for the identification of Sox10 targets, a high affinity antibody that recognises Sox10 would be required, something unavailable at present for zebrafish. Until recently suitable microarrays were not commercially available for zebrafish but this technique has recently been applied to zebrafish (Wardle et al., 2006) and the promoter chips designed for this paper are now available through Agilent Technologies. This technique can identify direct targets of the protein of interest while the microarray approach will identify both direct and indirect targets of *Sox10*. So while an extremely interesting and valuable technology, ChIP-on-chip was technologically too challenging to perform and did not have the same potential to identify a range of genes key to neural crest and ear development as the microarray approach presented in this chapter.

Another strategy that can be applied to identify the gene expression profile of a cell type and compare it to other expression profiles is Serial Analysis of Gene Expression (SAGE) and longSAGE. LongSAGE has been used to identify a molecular signature for Epidermal Neural Crest Stem Cells (EPI-NCSCs) (Hu et al., 2006). SAGE requires unique sequence identifiers (tags) from every mRNA molecule to be isolated, these are concatenated and sequenced. The number of times that a tag for a specific transcript appears is assumed to be relative to the abundance of that transcript and can therefore be used as a measure for gene expression. The relative abundance of transcripts can be compared between samples to identify differentially expressed genes. Again this technique could be applied to samples of cells sorted from transgenic embryos by FACS. A microarray approach was preferred to a SAGE approach as it is a more widely applied technique with many resources to support and facilitate data processing and analysis.

The microarray experiment presented here could be extended by extracting RNA from FACS purified GFP negative cells and comparing the gene expression profile generated from these cells to that generated from GFP positive cells (Covassin et al., 2006, Reeves and Posakony, 2005). This strategy can identify genes preferentially expressed in a cellular population of interest and is commonly used in combination with transgenic embryos. This is the first work we are aware of that combines FACS and microarray analysis to examine a mutant genotype. Using the multi class analysis option on GEPAS an analysis of variance test (ANOVA) or a CLEAR test could be applied to identify genes differentially regulated across WT GFP positive, *sox10*$^{-/-}$ GFP positive and

GFP negative cells from zebrafish embryos. It is possible that such an approach would facilitate the removal of false positives from the list of differentially regulated genes thus improving the hit rate. This approach was not pursued as the cost of Affymetrix GeneChips was a limiting factor. Collaborators are working on a similar project to examine differentially regulated genes in *mitfa*$^{-/-}$ embryos. This data could then be used to expand the project to include a comparison of two different mutants. This approach may facilitate the identification of melanocyte specific genes.

Chapter 4: Validation of *sox10* targets identified by Microarray analysis

4.1 Introduction

Lists of statistically significant differentially expressed genes between WT and *sox10*$^{-/-}$ embryos were generated from microarray data in the previous chapter. To validate the results from the microarray analysis a whole mount *in situ* hybridization approach was adopted. The *7.2sox10*:GFP data was gathered before the *4.9sox10*:GFP data and was therefore used to identify targets of *sox10* for validation. An unbiased approach was adopted to assess the microarray results; the most down-regulated genes and the most up-regulated genes ranked by *t*-statistic were cloned and *in situ* hybridization probes were generated. Gene expression was examined in WT and *sox10*$^{-/-}$ embryos at both 24 hpf and 30 hpf time points to facilitate identification of differentially regulated genes. A total of 136 genes (89 down-regulated genes and 47 up-regulated genes) were examined by *in situ* hybridization for NC or otic vesicle expression that was differentially regulated in *sox10* mutant embryos compared to WT embryos. Of the down-regulated targets, 25/89 (28 %) individual genes (30/89 or 34 % including duplicate probe sets) were validated, of the up-regulated genes 2 were validated (4 %) giving a total of 27 (20 %) validated differentially-regulated genes.

4.1.1 Methods that can be applied to validate Microarray Data

Microarray experiments generate a large amount of data regardless of the array type used or biological question being examined. The data presented in Chapter 3 indicates that hundreds of genes are differentially regulated in *sox10*$^{-/-}$ embryos. In fact the data sets generated will include results that are not biologically relevant as a consequence of mRNA levels varying between WT and mutant samples by chance alone and as a result of transcript to transcript variations in the efficacy of various technical steps. Additionally genes will have been determined incorrectly as differentially regulated as a result of weaknesses in the *t*-test, including a tendency to report genes as significant due to low variance or as a consequence of not controlling for multiple testing (Cui and Churchill, 2003). As such it is vital to confirm the biological significance of targets identified in a microarray screen. Two methods are typically utilised to validate microarray targets, real time quantitative or semi-quantitative RT-PCR and *in situ* hybridization. These methods are often used to screen a list of transcripts generated from microarray data for genes of interest not previously associated with the biological variable under scrutiny, novel genes of interest.

Semi-quantitative RT-PCR determines the ratio between the abundance of a target transcript relative to the amount of a control transcript, typically a house keeping gene. The expression level of this house keeping gene does not vary under the experimental conditions being tested. Microarray analysis was performed to identify genes differentially regulated in the zebrafish *cloche* mutants and semi-quantitative RT-PCR was performed to validate 19 novel potential down-regulated targets. Of the 19 targets tested, 13 showed clear down-regulation (Qian et al., 2005). In a similar study, again examining the *cloche* mutant, 24 genes of interest were examined by semi-quantitative RT-PCR and 13 were validated by this method (Sumanas et al., 2005). Quantitative RT-PCR enables the absolute amount of a transcript to be determined, typically expressed as a copy number or a concentration value. This is achieved using an external standard that has the same amplification efficiency as the target sequence of interest. Various different amounts of this standard are amplified and a standard curve of crossing point (point at which fluorescence and therefore cDNA levels are detectable above background) against log of the amount of standard is generated. The sample of interest can be amplified and by comparison to the standard curve the amount of starting material can be quantified. This technique was used to validate four genes identified as differentially expressed in the eye of mouse *Mitf*^{vitiligo} mutants by microarray analysis (Gelineau-van Waes et al., 2008). Both techniques are very sensitive and enable the quantification of mRNA levels using specific primers and this can corroborate data from microarray analysis. Additionally RT-PCR experiments can be performed in large batches enabling rapid high-throughput validation.

Real time RT-PCR is a powerful and sensitive quantitative technique but it is difficult to generate comprehensive data on spatial gene expression patterns using this method. Whole mount *in situ* hybridization is a qualitative technique that addresses these patterns of gene expression. Therefore genes identified by microarray analysis and validated by RT-PCR are often further validated and examined using *in situ* hybridization. Qian et al., 2005 examined the 13 genes previously validated by RT-PCR as differentially regulated in *cloche* mutants by *in situ* hybridization. Eight of these target genes showed expression in cell types of interest. The remaining five genes did not display a specific expression pattern. Sumanas et al., 2005 also examined eight of the validated genes by *in situ* hybridization. Seven of these genes were expressed in vascular endothelial cells. Other microarray studies have employed *in situ* hybridization alone to validate target genes. A microarray comparison of drosophila proneural cluster cells versus non-proneural cluster cells identified 204 genes preferentially expressed in the cells of interest. Subsequently 43 genes were analysed by *in situ* hybridization. 27 novel candidates were validated in this way which showed specific expression patterns in the cell types of interest (Reeves and Posakony, 2005). Typically validation and characterisation of genes by *in situ* hybridization only focuses on a handful of genes. Rarely, microarray studies

have been validated by more comprehensive *in situ* hybridization screens. A microarray study undertook an analysis of differential gene expression between two populations of *Xenopus* ectodermal cells, superficial polarised epithelial cells and deep non-epithelial cells. This study identified 98 differentially expressed genes, all of which were examined by *in situ* hybridization. From this second screen 32 differentially regulated genes were positively validated (Chalmers et al., 2006). Such large scale *in situ* hybridization screens are rare due to the labour intensive nature of generating a substantial number of *in situ* hybridization probes. Standalone large scale *in situ* hybridization screens have been performed, particularly in zebrafish (Thisse et al., 2004). Such screens have been collated on the zebrafish information network (ZFIN) website (www.zfin.org) (Sprague et al., 2006). These databases and other published data have been used where available to validate microarray data (Covassin et al., 2006, Cerda et al., 2009). This use of published data represents an easy way to examine the expression patterns of genes identified by microarray analysis. However these screens are not comprehensive and will not enable differential expression between WT and mutant embryos to be determined.

The methods described above can detect differential gene expression between two biological situations. Both real time RT-PCR and *in situ* hybridization are capable of being used in a high throughput manner to screen and validate a large number of genes that can, for example, be generated from microarray data. Real time RT-PCR presents a rapid method to validate candidate genes. However when applied to cells of interest purified by FACS, this technique would not add additional information, it would only corroborate the microarray data. Indeed RT-PCR would not help in differentiating between otic vesicle expressed genes and NC expressed genes. In contrast the less sensitive technique of whole mount *in situ* hybridization is capable of revealing information on both spatial and temporal gene expression patterns. In the case of NCCs, in which different cell types migrate on stereotypical pathways, this method would help to elucidate the cell types that novel candidate genes are expressed in.

4.1.2 Screening for Genes Important in Neural Crest and Otic Development

Chemical mutagenesis screens carried out in zebrafish have identified a large number of mutants with critical roles in vertebrate development (Haffter et al., 1996, Driever et al., 1996). Point mutations were induced in the germ cells of adult male zebrafish using *N*-ethyl *N*-nitrosourea (ENU), a chemical mutagen. Treated male zebrafish were crossed with untreated females to transmit the mutation and generate the F_1 generation. F_2 families were produced by crossing F_1 siblings, progeny from crosses between F_2 siblings were then examined for homozygous mutant embryos. Nearly 100 mutants from separate complementation groups affecting NC pigment cell development were phenotypically

characterised (Kelsh et al., 1996, Odenthal et al., 1996). Two studies focused on mutants that affected the development of the zebrafish ear, one study identified 39 mutants in separate complementation groups (Whitfield et al., 1996) while the second study identified 13 mutants (Malicki et al., 1996) although complementation tests were not performed. Therefore these screens identified a large number of mutants and provided insights into the biological processes underpinning the development of the NC (Kelsh, 2004) and the otic vesicle. Subsequently the cloning of some of these genes has been successful, for example the *colourless* mutant was successfully identified as a lesion in the zebrafish *sox10* gene (Dutton et al., 2001a). However the positional cloning of genes is a laborious process and many genes corresponding to characterised mutants have remained elusive. Regardless of this weakness, these chemical mutagenesis screens remain an important resource in vertebrate biology.

To address the difficulty of cloning a gene altered in a mutant of interest an insertional mutagenesis screen was undertaken in zebrafish to identify genes important in early development (Amsterdam et al., 2004). Insertional mutagenesis was performed using exogenous DNA as a mutagen, the DNA also acts as a molecular tag to facilitate the identification and cloning of the gene at the locus of the insertion event (Pickart et al., 2004). The insertional mutagenesis screen identified approximately 390 separate mutants and from these, 315 corresponding genes were identified and cloned (Amsterdam et al., 2004). This screen did not focus on genes important in NC or otic vesicle development but it included 11 mutants in which melanocyte pigmentation was altered, 9 of these 11 genes encoded subunits of Vacuolar-ATPase (V-ATPase) ion transporters, V-ATPase associated proteins or proteins associated with intracellular vesicles (Amsterdam et al., 2004). Thus while this resource did not focus on genes important in the NC, some were identified.

A combination of chick embryological techniques and array technology was applied to identify genes expressed in recently induced NC tissue (Gammill and Bronner-Fraser, 2002). Pre-migratory NCCs were induced *in vitro* by juxtaposing pieces of non-neural ectoderm and intermediate neural plate. Subsequently, cDNA was generated from induced NC tissue and from individual non-neural ectoderm and neural plate tissue samples for comparison using a macroarray. The macroarray was prepared using a cDNA library generated from 4 to 12 somite chick embryos, a period that covers early events in chick NC development and strong NC expression of *sox10* (Cheng et al., 2000). Genes that were up-regulated on the macroarray in induced NCCs in comparison to the non-NC tissue samples were identified. This highlighted the 97 genes with the greatest fold changes as being expressed as a result of NC induction; all of these genes were then examined by *in situ* hybridization and 83 genes were validated (Gammill and Bronner-Fraser, 2002). A follow up examination of the data generated from the study by Gammill

and Bronner-Fraser, 2002 selected all remaining up-regulated genes for *in situ* hybridization analysis (Adams et al., 2008). This additional analysis was performed to identify weakly expressed genes with the aim of finding key NC transcription factors. Over 300 genes were identified and 112 genes were examined by *in situ* hybridization leading to the validation of 101 NC genes including 14 transcription factors (Adams et al., 2008). The transcription factor *sox10* was not identified as up-regulated in NC compared to non-NC tissue by either of these studies. With the NC being a difficult tissue to dissect from embryos this *in vitro* approach has helped to identify a large number of NC expressed genes but may not fully recapitulate the true *in vivo* expression profile of recently induced NCCs.

Multiple types of neural crest stem cell (NCSC) have been discovered both in developing embryos and adult organisms (Delfino-Machín et al., 2007). The different types of NCSC studied display a wide range of different markers and these have been reviewed by Delfino-Machín et al., 2007. While a common NCSC molecular signature has not been identified, a molecular signature for embryonic NCSCs and epidermal NCSCs (EPI-NCSCs) has been characterised (Hu et al., 2006). The transcriptome of NCSCs, EPI-NCSCs and differentiated NCCs were captured by LongSAGE analysis and then compared. Genes abundant in NCSCs and in EPI-NCSCs but not in differentiated NC progeny were identified giving a characteristic gene expression profile of 91 genes. To generate the NCSC molecular signature, genes also expressed in bulge epidermal cells of the hair follicle were eliminated leaving a profile of 19 genes (Hu et al., 2006). Melanocyte stem cells (MSCs) have also been characterised molecularly in their niche, the lower permanent portion of the hair follicle (Osawa et al., 2005). MSCs were characterised by PCR based gene expression profiling of individual genetically labelled microdissected cells. Strikingly MSCs were mainly characterised by the genes that they did not express but conversely were highly expressed by melanocytes and to a lesser extent melanoblasts (Osawa et al., 2005).

A further set of screens that have identified a large number of NC expressed genes have already been alluded to, the *in situ* hybridization screens performed on whole mount zebrafish that are collected on the ZFIN website. Additionally published zebrafish expression patterns are also collated into the ZFIN database giving an extensive searchable resource. A search of the gene expression database on ZFIN at the time of writing using the anatomical term neural crest returned 219 genes with expression patterns (Sprague et al., 2006).

Most approaches have attempted to identify the entire transcriptome for the cell type of interest. Complimentary to this would be the identification of the proteome of a cell type of interest. One study has endeavoured to elucidate the proteome of early

melanosomes purified from pigmented human MNT1 cells (Basrur et al., 2003). A total of 68 proteins were identified from early melanosomes including 6 known proteins that are specific to melanosomes. By focusing on a melanocyte specific organelle this study has unearthed a sizeable number of genes likely to play a key role in melanosome and therefore melanocyte biology.

4.1.3 Aims

This chapter aims to validate the genes identified from the *7.2sox10*:GFP microarray analysis as differentially regulated in *sox10*$^{-/-}$ embryos when compared to WT embryos. Genes were validated by performing an *in situ* hybridization screen. Validated genes were expressed in the NC or otic vesicle of zebrafish and were differentially expressed in mutant embryos compared to WT embryos. This approach was designed to generate information on both the spatial and temporal expression patterns of these genes.

4.2 Results and Conclusions

4.2.1 Description of the *in situ* Hybridization Screen

An *in situ* hybridization screen approach was selected to validate the candidate transcripts of interest identified in the *sox10* mutant *7.2sox10*:GFP microarray screen. A blind screen was initiated to validate the most differentially regulated genes, with the aim being to screen as many genes as possible. Genes were screened from the 100 most down-regulated and 50 most up-regulated candidates as ranked by *t*-statistic. Affymetrix Identifiers were annotated with available gene symbol information as supplied on the NetAffx™ Analysis Centre website (www.affymetrix.com/analysis/index.affx). To generate *in situ* hybridization probes complimentary to each candidate transcript, the following pipeline was employed.

Sequence data was obtained for every candidate as a preliminary step to *in situ* hybridization probe design. Sequence data for a well characterised transcript was taken from the NCBI website (www.ncbi.nlm.nih.gov) Entrez Nucleotide database by referring to the transcript assignment provided by Affymetrix. In the majority of cases sequence data was obtained in this way. When annotation was poorly defined, with multiple transcript assignments or given an Affymetrix determined poor annotation grade (grade E or R), target sequence data supplied on the NetAffx™ Analysis Centre website was utilised. On occasions a BLAST search on either the ZFIN website or the NCBI website using the Affymetrix supplied target sequence data yielded additional sequence data. Primers were then designed to amplify candidate sequences from 24 hpf zebrafish cDNA. Typically

primers were designed to amplify a 500 to 600 bp fragment of the transcript of interest. In circumstances when annotation was poor and only a short fragment of sequence data was available the largest possible fragment of sequence was targeted for amplification.

Amplified PCR fragments were gel extracted, if purified DNA fragments were of a very low concentration a second identical PCR step was implemented to amplify the amount of purified DNA and then gel extraction was repeated. All candidates were then cloned into a pGEMT®-Easy vector. After transformation using competent *E. coli*, blue/white colony selection was performed on X-gal supplemented agar plates to identify bacteria containing plasmid with an insert. Plasmid DNA was purified from liquid bacteria cultures by miniprep. A diagnostic restriction enzyme digest using EcoRI was performed to check for the presence and the size of the DNA inserted into the plasmid. Plasmids were then sequenced to check that the correct fragment had been ligated into the vector and in what orientation. This information was exploited to determine which RNA transcription start site present in the pGEMT®-Easy vector (T7 or SP6) should be used for generating a labelled anti-sense *in situ* hybridization probe. After synthesising probes, the expression pattern of each candidate was determined using fixed 24 hpf and 30 hpf WT and *sox10*$^{-/-}$ embryos. Embryos at 24 hpf were used as this was the time point of interest, 30 hpf embryos were used as pigmentation is visible at this stage to facilitate the genotyping of embryos. The embryos were examined to identify NC and otic vesicle expression and any differences between WT and mutant embryos were determined.

4.2.2 Generating an *in situ* hybridization probe for Dr.13165.1.A1_at

Results from the steps performed to synthesise the *in situ* hybridization probe for candidate Dr.13165.1.A1_at are presented in this section to demonstrate the pipeline that was followed. Dr.13165.1.A1_at was annotated on the NetAffx™ Analysis Centre website as si:dkey-199l1.2 and sequence data was obtained from the NCBI website. Originally this candidate was annotated as *npr3* however this annotation was removed but the sequence corresponding to this candidate did not change. A left primer (ACGAGAGTTTTTCCGGCTAT) and right primer (TACTATGTCCGCCTTTGAGC) were designed to amplify a 570 bp fragment of the si:dkey-199l1.2 transcript. PCR was performed to amplify the transcript of interest from 24 hpf zebrafish cDNA (**Figure 16A**). The DNA fragment subsequently purified from this gel by gel extraction is shown in **Figure 16B**. After ligating the fragment into a pGEMT®-Easy vector and transforming the plasmid into bacterial cells, plasmid DNA was purified by miniprep (**Figure 16C**). Confirmation that purified plasmid DNA contained an insert of the correct size was obtained through a restriction enzyme digest, an insert of approximately 570 bp was released from the plasmid DNA (**Figure 16D**).

Once the presence of insert was confirmed, plasmid DNA was sent for sequencing using T7 sequencing primers and the results analysed (**Figure 17**). Both primers were identified within the sequence data demarcating a region of 570 bp. A BLAST search with this sequence returned a result of 99 % similarity with si:dkey-199l1.2. Note that the BLAST search indentified the sequence data as being in the reverse compliment (strand minus/plus) orientation. Thus the correct gene sequence has been cloned in a 3' to 5' prime direction from the T7 site, or in the 5' to 3' direction from the SP6 site. To generate an anti-sense *in situ* hybridization probe, transcription must start from the 3' end thus T7 polymerase was used for probe synthesis. The SalI (Sal1) restriction enzyme site 5' of the insert, towards the SP6 site, a site not present within the insert sequence, was selected to linearise the plasmid.

An antisense *in situ* hybridization probe was synthesised as shown above and whole mount *in situ* hybridization was performed on 24 hpf and 30 hpf zebrafish embryos (**Figure 18**) from a cross that generated 25 % homozygous *sox10* mutants. According to ZFIN, si:dkey-199l1.2 is an old name for the gene *npr3* and directly submitted gene expression images show expression in the floor plate, hypochord, neural tube and pronephric ducts (Thisse et al., 2004). These results show strong similarities to the expression pattern shown in **Figure 18**. In addition to the published sites of expression it is noted that scattered neural tube cells positioned just above the floor plate also express *npr3* (**Figure 18A**). No obvious expression was detected in the NC or otic vesicle and no differences were detected between the expression pattern of *npr3* at 24 hpf and 30 hpf or between WT and *sox10*$^{-/-}$ embryos (data not shown). Therefore, *npr3* is a false positive result. These results show that the pipeline adopted to generate the *in situ* hybridization probes for the screen can produce probes with very strong and specific signals that perform at least as well as published probes and which allow testing of the validity of the microarray data.

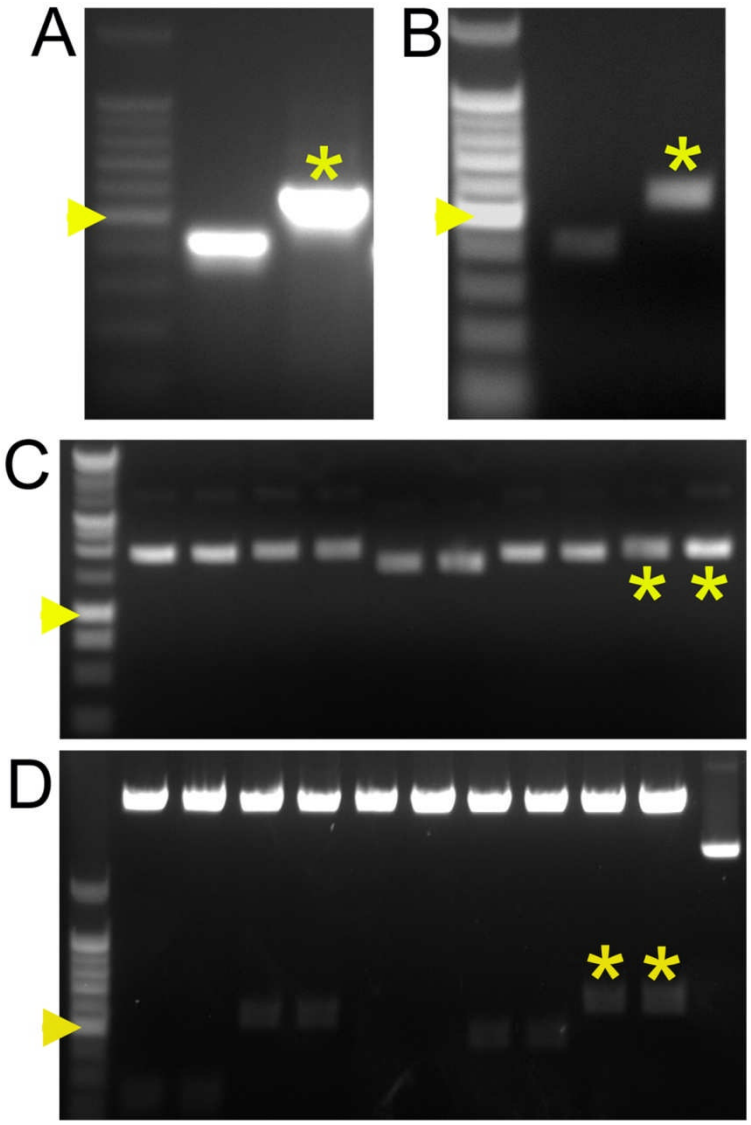


Figure 16: Early Steps during *in situ* Hybridization probe synthesis.
All images are of ethidium bromide stained agarose gels. The size marker used in gels A, B and D was a 100 bp ladder (Promega), the 500 bp size marker is indicated by a yellow arrowhead. The size marker used in gel C was a 1 Kb ladder (Promega) with the 1,000 bp size marker indicated by a yellow arrowhead. Bands pertaining to the transcript of interest (si:dkey-199l1.2) are indicated with an asterisk (*), unlabelled bands are from other samples at the same stage in the pipeline. The expected band size of si:dkey-199l1.2 was 570 bp. (A) A band of approximately the expected size for si:dkey-199l1.2 was amplified by PCR. (B) The same band of amplified DNA in A is shown after gel purification. (C) Two purified plasmid DNA samples that should contain the si:dkey-199l1.2 fragment have been visualised. (D) An EcoRI restriction digest was performed on the plasmid DNA samples shown in C, asterisks highlight the inserts released from the plasmid DNA that are approximately 570 bp in size. The lane on the far right of (D) shows undigested plasmid DNA while the bright band at the top of the gel picture in all other lanes represents linear plasmid DNA.

> Dr.13165.1.A1_at

ATGCTCCGGCCGCCATGGCGGCCGCGGGAATTCGATTTACTATGTCCGCCTTTGAG
CACATGATGACCACTTCGCTCCCGTAGACGGAGGTGATGATCCCGTCGGGGTCTAC
CCTCTCCTCGTTGGAGTTCAACACGGCAAAGTCTGTTGAGATGTGGTACTCAGAGAG
CACCGTAAAAACTCCCTCCATCGTAAAGTAACAGTTTCTCTCATCTTTATCATCGTCG
TAAATCAAATACGCGGTTCTCCAGCCGAAATGTCCAAATATCGCCTGAAAGGTCTCA
GCCATCTTCAGGTATGTTGGTGCGATTCGTGTCAGATGAGAGTACTCGGGGGTTTTG
CTGTTGAAGCCGGTGGCCAGAGCACCTGCCGAGATCACCGGGATGTTCCAGTGCGA
TGCCACCCTCGTGACCGACGAAGCCGCGTACTCGCACACCGGACCGAGGACCAGG
TCGGGACGCTCGTCTTTCTGCCGGTCCACGAGCGCGTAGAGAGCGTCCATCCCGCA
CGCGGAGTTCTCAAAGCGCACATTGAATCTCAGTCCCGCATACGCGTCTGCTTCTCT
GAGGGCTTTTTTCGCGTACTCAATAGCCGGAAAAACTCTCGTAATCACTAGTGAATTC
GCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCATAGC
TTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTC
CTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAA
GTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTC
ACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCA
ACGCGCGGGGAGAGGCGGTTTGCGTAT

BLAST search results: Identities = 567/570 (99%), Positives = 567/570 (99%), Strand = Minus / Plus to Danio rerio si:dkey-199l1.2, mRNA

Figure 17: Dr.13165.1.A1_at Sequence Data.
Both primers were located in the returned sequence data, the right primer appears first and the reverse compliment of the left primer appears second, both highlighted in green. The results of a ZFIN BLAST search against RNA and cDNA sequences using the sequence between both primers was performed with the sequence data returning a hit of 99 % similarity to si:dkey-199l1.2.

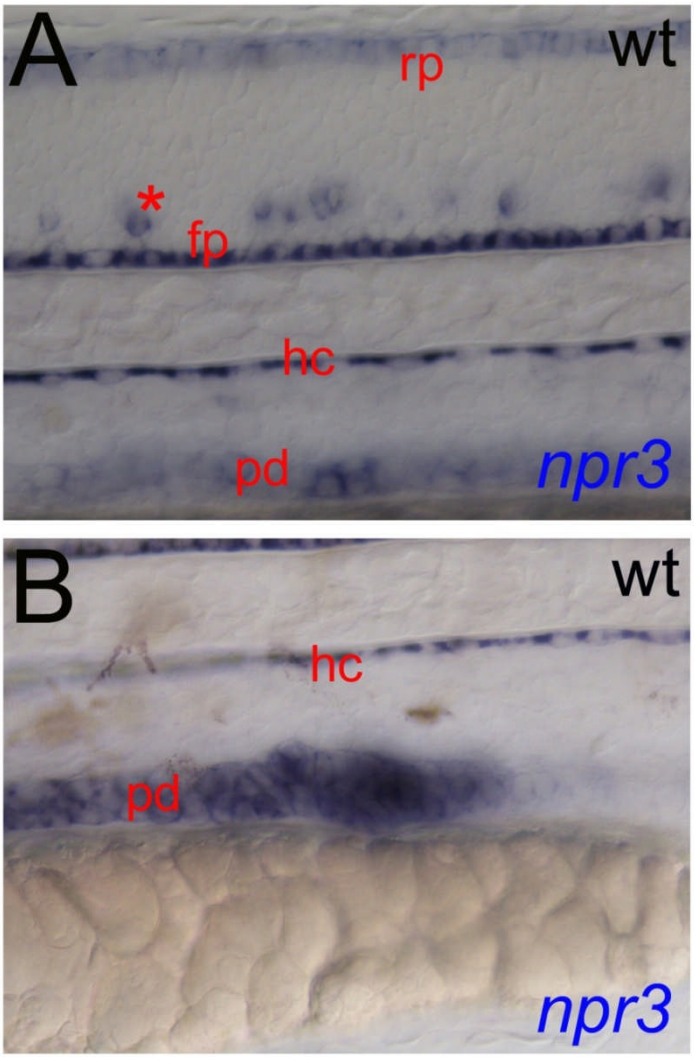


Figure 18: Whole mount *in situ* hybridization pattern of *npr3* expression.
Lateral views of 30 hpf zebrafish trunks are shown oriented, dorsal top, ventral bottom, anterior left and posterior right. An image of mid-trunk (A) and an image of a more posterior trunk region towards the end of the yolk sac extension (B) are shown. Expression of npr3 occurs in distinct stripes stretching the entire length of the trunk, rp = roof plate, fp = floor plate hc = hypochord and pd = pronephric duct. Pronephric duct expression is stronger posteriorly (B). In addition to expression in the labelled structures, npr3 is also expressed in scattered cells in the ventral neural tube indicated by an asterisk (*) in A.

4.2.3 Summary of the Down-Regulated Genes Selected for Validation

The aim was to validate as many genes as possible from the 100 most down-regulated candidates, ranked by *t*-statistic (**Table 15**). Seven transcripts were not examined as a result of time constraints, these are summarised in **Table 16**. Of the seven candidates not examined, five could not be amplified from 24 hpf cDNA, one repeatedly failed to ligate successfully into a vector and for one, primers were not designed. From the remaining 93 candidates there were three pairs of probe sets that corresponded to the same genes (**Table 17**). Thus when duplicates were ignored, 90 separate genes were investigated. Initial annotation identified the probe Dr.24498.1.A1_at as *sox11a* and as such sequence data pertaining to this gene was used to generate an *in situ* hybridization probe. Subsequent to the screen being completed this annotation was removed and as such the wrong gene was cloned and examined. Therefore 89 separate candidates were successfully examined, included within this list were five marker genes that were expected to be down-regulated (*crestin, dct, foxD3, mitfa* and *otomp*). Plasmids used for the transcription of *in situ* hybridization probes to detect mRNA transcribed from these genes were already available in the laboratory and as such were not cloned. The *otomp* probe was obtained from Dr. Tanya Whitfield. The *pax7* probe was obtained courtesy of Dr. Hitoshi Okamoto. Thus for the *in situ* hybridization screen 83 candidates were cloned and 89 gene expression patterns examined in WT and *sox10*$^{-/-}$ embryos at two separate time points. Whole mount *in situ* hybridization expression patterns are shown for the validated genes. These are ordered into sections based on the cell type they are likely to be expressed in.

Affymetrix ID	*t*-statistic	*p*-value	Fold Change	Gene Symbol
Dr.991.2.A1_s_at	-6.173	0.003497	-1.15	*ggps1*
Dr.22144.1.A1_at	-6.185	0.003473	-1.25	*zgc:113142*
Dr.7154.1.A1_at	-6.227	0.003386	-1.19	*slc25a26*
Dr.19421.1.A1_at	-6.231	0.003378	-1.24	*phf8*
Dr.1251.1.S1_at	-6.27	0.003303	-1.58	*degs1*
Dr.7955.1.A1_at	-6.274	0.003295	-1.5	
Dr.10151.1.A1_a_at	-6.282	0.003279	-1.37	
Dr.24473.1.A1_at	-6.296	0.003252	-1.11	*si:ch211-218c6.1*
Dr.14787.1.A1_at	-6.307	0.003231	-2.85	*s100b*
Dr.9236.2.A1_at	-6.365	0.003124	-1.45	*si:ch211-243g18.2*
Dr.11740.1.A1_at	-6.4	0.00306	-1.08	*wu:fb74h01*
Dr.1242.1.S1_at	-6.401	0.003059	-1.38	*cotl1*
Dr.5325.1.A1_at	-6.428	0.003011	-2.27	*zgc:65870*
Dr.21407.1.A1_at	-6.482	0.00292	-1.16	
Dr.13966.1.S1_at	-6.522	0.002854	-1.29	*zgc:63767*
Dr.6142.1.A1_at	-6.633	0.002681	-2.85	*si:dkey-180p18.9*
Dr.19416.1.S1_at	-6.638	0.002674	-1.86	*otomp*
Dr.7966.1.S1_at	-6.639	0.002672	-1.84	*atp6v1e1*
Dr.7250.1.A1_at	-6.676	0.002617	-1.41	*zgc:110239*
Dr.3991.1.A1_at	-6.692	0.002594	-2.53	*wu:fd02h10*

Dr.20271.1.S1_at	-6.705	0.002575	-1.14	*zgc:162119*
Dr.4599.1.A1_at	-6.854	0.002373	-1.29	*zgc:112247*
Dr.16013.1.A1_at	-6.873	0.002348	-1.29	*wu:fj65e12* *LOC100004591*
Dr.4751.2.A1_at	-6.883	0.002335	-2.48	*aldh2b*
Dr.7215.1.S1_at	-6.947	0.002256	-1.25	*mtch2*
Dr.4615.2.S1_at	-6.975	0.002222	-1.06	
Dr.22212.1.A1_at	-6.98	0.002216	-1.24	*cyp3c1l2*
Dr.10336.1.S1_at	-6.987	0.002207	-9.59	*dct*
Dr.1920.1.S1_at	-7.018	0.002171	1.07	*derl1*
Dr.13359.1.S1_at	-7.087	0.002093	-1.84	*adsl*
Dr.9157.1.S1_at	-7.126	0.00205	-1.3	*mkks*
Dr.17743.1.A1_at	-7.173	0.002	-1.94	*zgc:66482*
Dr.13642.1.S1_at	-7.181	0.001992	-1.91	
Dr.14845.1.S1_at	-7.359	0.001816	-1.1	
Dr.20896.1.S1_at	-7.372	0.001805	-1.99	*slc2a15b*
Dr.4171.1.A1_at	-7.446	0.001738	-3.7	*pah*
Dr.7590.1.A1_at	-7.478	0.00171	-2.02	*wu:fj68b05*
Dr.4248.1.A1_at	-7.673	0.001551	-1.52	*wu:fa98e08*
Dr.10624.2.S1_a_at	-7.705	0.001527	-2.84	*zgc:110343*
Dr.18158.1.A1_at	-7.71	0.001523	-1.53	
Dr.8124.1.S1_at	-7.728	0.00151	-2.49	*crestin*
Dr.4751.1.S1_a_at	-7.748	0.001495	-2.34	*aldh2a* *aldh2b*
Dr.983.1.S1_at	-7.758	0.001488	-1.56	*paics*
Dr.26101.1.A1_at	-7.819	0.001444	-3.26	
Dr.4629.1.A1_at	-7.83	0.001436	-1.53	*pi4k2a*
Dr.22990.1.A1_at	-7.88	0.001402	-3.05	*wu:fj82h06*
Dr.1691.3.S1_at	-7.89	0.001396	-1.4	*cct5*
Dr.18052.1.S1_at	-7.898	0.00139	-1.4	*znf593*
Dr.8594.1.S1_at	-7.945	0.001359	-1.81	*zgc:100919*
Dr.3237.1.A1_at	-7.946	0.001358	-1.57	*wu:fc08b04*
Dr.13244.1.A1_at	-7.965	0.001346	-1.57	*zgc:152987*
Dr.26323.1.A1_at	-7.965	0.001346	-1.35	*wu:fb95f11*
Dr.8080.1.S1_at	-8.03	0.001305	-2.14	*mitfa*
Dr.6806.1.S1_at	-8.161	0.001227	-2.8	*keap1*
Dr.10624.2.S1_at	-8.217	0.001195	-2.08	*zgc:110343*
Dr.3972.1.S1_at	-8.302	0.00115	-3.2	
Dr.2155.2.S1_at	-8.385	0.001107	-1.22	*ppa1*
Dr.590.1.S1_at	-8.529	0.001037	-1.32	*foxd3*
Dr.25954.1.A1_at	-8.543	0.001031	-1.2	
Dr.8251.1.A1_at	-8.556	0.001025	-1.6	*wu:fi05a09*
Dr.481.1.S1_at	-8.667	0.000975	-1.85	*atic*
Dr.12707.1.A1_at	-8.672	0.000973	-2.16	*zgc:112072*
DrAffx.1.65.S1_at	-8.688	0.000966	-1.54	*cx33.8*
Dr.2606.1.A1_at	-8.689	0.000966	-1.26	*LOC568088*
Dr.2788.1.S1_at	-8.714	0.000955	-2.63	*si:dkeyp-86b9.2*
Dr.2393.1.S1_at	-8.766	0.000934	-1.14	*dohh*
Dr.14798.1.A1_at	-8.793	0.000923	-1.19	*zgc:110718*
Dr.2729.1.A1_at	-8.801	0.000919	-1.57	*wu:fc54b10*
Dr.4873.1.A1_at	-8.871	0.000892	-1.16	*wu:fi46h04*
Dr.2901.1.S1_at	-9.323	0.000737	-1.45	*coro1c*
Dr.646.1.S1_at	-9.401	0.000713	-1.26	*mg:cb01g09*
Dr.15700.1.A1_at	-9.492	0.000687	-1.31	*ell*
Dr.4568.1.A1_a_at	-9.55	0.000672	-1.19	*wu:fc17h11*

Dr.3180.1.A1_at	-9.713	0.000629	-1.96	*bhlhb2*
Dr.20850.1.S1_at	-9.913	0.000581	-1.69	*fabp7a*
Dr.18800.1.S1_at	-9.97	0.000569	-1.27	*ccbl2*
Dr.14747.1.A1_at	-9.988	0.000565	-3.03	*slc2a15b*
Dr.7249.1.A1_at	-10.191	0.000522	-1.14	*zgc:110397*
Dr.5963.1.A1_at	-10.548	0.000457	-1.62	*wu:fj38e08*
Dr.23456.1.S1_at	-10.592	0.00045	-1.25	*pax7*
Dr.22614.1.A1_at	-10.704	0.000432	-3.22	
Dr.24498.1.A1_at	-10.723	0.000429	-1.28	
Dr.4838.1.A1_at	-10.73	0.000428	-3.36	*wu:fb51g06, zgc:158673*
Dr.13012.1.S1_at	-10.857	0.000408	-1.66	*nfkb2*
Dr.3933.1.A1_at	-11.345	0.000344	-3.58	*wu:fc46b01, zgc:100933*
Dr.25879.1.A1_at	-11.355	0.000343	-1.13	
Dr.18194.1.A1_at	-11.487	0.000328	-2.62	*pik3c3*
Dr.2900.1.A1_at	-11.918	0.000284	-1.46	*ift88*
Dr.3154.1.A1_at	-12.091	0.000268	-1.29	*LOC564527 zgc:154036*
Dr.12107.1.A1_at	-12.15	0.000263	-2.14	*ndrg1*
Dr.25257.1.A1_at	-14.48	0.000132	-7.41	*bhlhb3l*
Dr.9611.1.A1_at	-14.789	0.000122	-3.11	*si:ch211-241e15.2 wu:fj94a03*
Dr.518.1.A1_at	-15.54	0.0001	-4.59	*sb:cb319*
Dr.10292.1.S1_at	-16.037	0.000088	-7.77	*rbp4l*
Dr.8680.1.S1_at	-16.124	0.000087	-2.71	*cldnj*
Dr.2855.1.A1_a_at	-16.267	0.000084	-1.86	*LOC564804 similar to ATPase, H+ transporting, lysosomal 70kDa, V1 subunit A, like*
Dr.9560.1.A1_at	-16.778	0.000074	-1.52	*wu:fj87g04 LOC565706*
Dr.13919.1.S1_at	-17.892	0.000057	-3.8	
Dr.4612.1.A1_at	-18.127	0.000054	-2.83	*wu:fc31e04*
Dr.16789.1.A1_at	-25.393	0.000014	-1.96	
Affymetrix ID	***t*-statistic**	***p*-value**	**Fold Change**	**Gene Symbol**

Table 15: Table of the 100 most down-regulated genes in *sox10* mutant embryos, ranked by *t*-statistic.

The 100 most down-regulated genes as identified from the *7.2sox10*:GFP microarray data analysed in Chapter 3 and evaluated using a *t*-Test are shown. Genes are ranked by *t*-statistic; the *p*-value generated by the *t*-Test is also shown. The fold change for each gene is also shown as determined from the CLEAR test. Each gene was annotated with a gene symbol using the NetAffx website.

Affymetrix Identifier	Gene Symbol	Reason That Gene Expression Pattern Was Not Obtained
Dr.9560.1.A1_at	wu:fj87g04, LOC565706	PCR failed
Dr.22614.1.A1_at		PCR failed
Dr.4568.1.A1_a_at	wu:fc17h11	Ligation failed
Dr.4615.2.S1_at		PCR failed
Dr.6142.1.A1_at	si:dkey-180p18.9	Primers not designed
Dr.21407.1.A1_at		PCR failed
Dr.14787.1.A1_at	*s100b*	PCR failed

Table 16: Summary of the down-regulated candidate genes for which assessment by *in situ* hybridization was not completed.
Candidates are identified by both Affymetrix identifier and gene symbol as annotated on the NetAffx™ Analysis Centre website. PCR failed indicates that the gene could not be amplified from 24 hpf zebrafish cDNA using the designed primers.

Affymetrix Identifier	Gene Symbol
Dr.14747.1.A1_at	*slc2a15b*
Dr.20896.1.S1_at	*slc2a15b*
Dr.10624.2.S1_a_at	*zgc:110343*
Dr.10624.2.S1_at	*zgc:110343*
Dr.4751.1.S1_a_at	*aldh2a* *aldh2b*
Dr.4751.2.A1_at	*aldh2b*

Table 17: Duplicate Down-Regulated Genes.

4.2.4 Summary of the Up-Regulated Genes Selected for Validation

It was attempted to clone and identify a gene expression pattern for as many of the up-regulated genes as determined by microarray analysis as possible. The 50 most up-regulated genes as ranked by *t*-test from the *7.2sox10*:GFP microarray data were examined (**Table 18**). From these 50 genes, 47 were successfully cloned and a gene expression pattern obtained, 3 genes were not cloned (**Table 19**). Within the 50 most up-regulated genes an Affymetrix control probe (AFFX-r2-Bs-lys-3_at) was present, this transcript was excluded thus the candidates that were screened actually comprised of 50 of the top 51 up-regulated genes. Although 47 of the up-regulated genes were cloned, the following probes were also obtained from other laboratories, *hoxd4a* with thanks to Dr. Victoria Prince and *wnt11r* with thanks to Dr. Simon Hughes.

Affymetrix ID	*t*-statistic	*p*-value	Fold Change	Gene Symbol
Dr.16.1.S1_at	14.94	0.00012	1.38	*nr2f1*
Dr.22087.1.A1_at	13.055	0.0002	1.47	*wu:fd12f01*
Dr.26266.1.A1_at	12.811	0.00021	1.2	
Dr.597.1.S2_at	11.376	0.00034	1.4	*vegfaa*
Dr.2967.1.A1_at	9.233	0.00076	1.95	*hoxc8a*
Dr.11268.1.A1_at	8.818	0.00091	1.71	
Dr.18334.1.A1_at	8.794	0.00092	1.25	*cars*
Dr.6063.1.A1_at	8.624	0.00099	1.58	*si:dkey-73n10.1*
Dr.21429.1.S1_at	8.527	0.00104	1.31	*zgc:85683*
Dr.23153.1.S1_at	8.394	0.0011	1.36	*LOC100000500*
Dr.17602.1.A1_at	8.277	0.00116	1.11	
Dr.4119.2.S1_at	8.189	0.00121	1.44	*hoxd4a*
Dr.22671.1.A1_at	8.166	0.00122	1.41	*si:dkey-174n20.1*
Dr.14080.1.A1_at	7.987	0.00133	1.28	*wu:fq26c12*
Dr.18932.1.A1_at	7.962	0.00135	1.08	*zgc:153973*
Dr.8297.1.S1_at	7.61	0.0016	1.59	*wnt11r*
Dr.4978.1.A1_at	7.513	0.00168	1.31	*zgc:85829*
Dr.1094.1.A1_at	7.334	0.00184	2.04	*wu:fk49e12*
Dr.11996.1.A1_at	7.316	0.00186	1.25	*LOC100000439*
Dr.12764.1.A1_at	7.288	0.00188	1.42	*sb:cb54*
Dr.3936.1.A1_at	7.177	0.002	1.43	*zgc:162730*
Dr.12263.1.A1_at	7.156	0.00202	1.39	*LOC556968*
Dr.20715.1.S1_at	7.143	0.00203	1.23	*ssr1*
Dr.10232.1.A1_at	7.1	0.00208	1.54	*meis2.2*
Dr.23461.1.A1_at	7.036	0.00215	1.53	*stanniocalcin 2 (stc2)* *zgc:136650*
Dr.1837.1.A1_at	6.988	0.00221	1.16	*wu:fc83f05*
Dr.17738.1.S1_at	6.915	0.00229	1.15	*tnfaip6*
Dr.14522.1.S1_at	6.862	0.00236	1.92	*pvalb7*
Dr.11252.1.A1_at	6.833	0.0024	1.1	*zgc:56085*
Dr.10119.1.A1_at	6.713	0.00256	1.16	*zgc:92480*
Dr.38.1.A1_at	6.7	0.00258	1.43	*sb:cb307*
Dr.18513.2.S1_at	6.617	0.0027	1.33	*sccpdhb*
Dr.7908.1.S2_at	6.589	0.00275	1.22	*calm1a* *calm1b* *calm2a* *calm2b* *calm3a* *calm3b* *zgc:55813*
Dr.21012.1.A1_at	6.478	0.00293	1.38	*sfrp5*
Dr.14529.1.S1_at	6.458	0.00296	1.19	*zgc:91935, LOC792266*

Dr.6496.1.S1_at	6.452	0.00297	1.18	*fth1*
AFFX-r2-Bs-lys-3_at	6.445	0.00298	2.63	
Dr.3336.1.S1_at	6.435	0.003	1.31	*meis1*
Dr.10254.1.A1_at	6.413	0.00304	1.37	*lin7b*
Dr.18367.1.A1_at	6.393	0.00307	1.19	
Dr.25500.1.S1_at	6.359	0.00314	1.27	*vamp1*
Dr.8093.1.S1_at	6.345	0.00316	1.25	*par1*
Dr.2615.1.S1_at	6.334	0.00318	1.34	*appb*
Dr.15532.1.A1_at	6.306	0.00323	1.13	*LOC100148646*
Dr.13165.1.A1_at	6.02	0.00384	2.01	*si:dkey-199l1.2*
Dr.18374.1.A1_at	5.961	0.00398	1.72	
Dr.16719.1.A1_at	5.957	0.00399	1.37	
Dr.1911.1.A1_at	5.927	0.00406	1.17	*LOC564069*
Dr.15347.1.S1_at	5.919	0.00408	1.42	*LOC562530*
Dr.10524.1.S1_at	5.903	0.00412	1.18	*zgc:85694*
Dr.4649.1.A1_a_at	5.777	0.00446	1.31	*phc2*
Dr.16431.1.S1_at	5.679	0.00474	1.24	*wu:fj63d08*
Affymetrix ID	***t*-statistic**	***p*-value**	**Fold Change**	**Gene Symbol**

Table 18: Table of the 50 most up-regulated genes in *sox10* mutant embryos, ranked by *t*-statistic.

The 50 most up-regulated genes as identified from the *7.2sox10*:GFP microarray data analysed in Chapter 3 and evaluated using a *t*-Test are shown. Genes are ranked by *t*-statistic; the *p*-value generated by the *t*-Test is also shown. The fold change for each gene is also shown as determined from the CLEAR test. Each gene was annotated with a gene symbol using the NetAffx website.

Affymetrix Identifier	Gene Symbol	Reason That Gene Expression Pattern Was Not Obtained
Dr.11996.1.A1_at	LOC100000439	PCR Failed
Dr.12764.1.A1_at	sb:cb54	Ligation Failed
Dr.18367.1.A1_at		PCR Failed

Table 19: Summary of the Up-Regulated Candidate Genes Not Examined during the *in situ* hybridization screen.

Candidates are identified by both Affymetrix identifier and gene symbol as annotated on the NetAffx™ Analysis Centre website. PCR failed indicates that the gene could not be amplified from 24 hpf zebrafish cDNA using the designed primers.

4.2.5 Validated Genes Down-Regulated in *sox10* mutant embryos.

4.2.5.1 Known down-regulated genes

The TF *mitfa* and the differentiation gene *dct* are both expressed in melanocytes and are known to be down-regulated in *sox10* mutant embryos (Dutton et al., 2001a, Kelsh and Eisen, 2000). In fact, evidence exists showing that both genes are direct targets of Sox10 (Bondurand et al., 2000, Ludwig et al., 2004). Thus the results from the microarray data and the *in situ* hybridization screen that both these genes were strongly down-regulated in NC derived melanoblasts at 24 hpf embryos were expected (**Figure 19** and **Table 20**). Occasional *dct* expressing cells can be seen in *sox10* mutant embryos thus expression of this differentiation gene is not completely abolished (Kelsh et al., 2000b).

The Affymetrix probe set Dr.8124.1.S1_at corresponds to the pan-NC marker gene *crestin* (Luo et al., 2001, Rubinstein et al., 2000). *Crestin* is a widely used NC marker but this gene has been poorly characterised. It has been hypothesised that *crestin* is a member of a family of retroelements in the zebrafish genome (Rubinstein et al., 2000). It has been shown previously that *crestin* is down-regulated in *sox10* mutant embryos (Elworthy et al., 2005). Note that in contrast to *dct* and *mitfa*, *crestin* expression is only reduced not absent. For example, a small number of cells expressing *crestin* can still be detected migrating on the medial pathway in mutant embryos (**Figure 19** and **Table 20**). These cells are likely to be of a neural fate and can be visualised by *sox10 in situ* hybridization in *sox10* mutant embryos (Dutton et al., 2001a). Thus the results from the microarray data and the *in situ* hybridization screen that *crestin* is strongly down-regulated in 24 hpf embryos was expected (**Figure 19**).

Affymetrix Probe Identifier		Dr.8080.1.S1_at	Dr.10336.1.S1_at	Dr.8124.1.S1_at
Affymetrix Gene Symbol		*mitfa*	*dct*	*crestin*
ZFIN Gene Symbol		*mitfa*	*dct*	*crestin*
Full Name		*microphthalmia-associated transcription factor a*	*dopachrome tautomerase*	*crestin*
Orthology		*MITF*	*DCT*	N/A
Biological Process GO Term		transcription, DNA-dependent melanocyte differentiation regulation of transcription	metabolic process	N/A
Molecular Function GO Term		transcription activator activity protein dimerization activity	oxidoreductase activity metal ion binding	N/A
7.2sox10:GFP array results	*t*-statistic	-8.0299043655	-6.9873137474	-7.7277312279
	p-value	0.0013052389	0.0022070811	0.0015099111
	CLEAR result	diff.expr	diff.expr.high.var	diff.expr
	Fold Change	-2.14	-9.59	-2.49
4.9sox10:GFP array results	*t*-statistic	-3.85014081	-5.4718756676	-3.1706631184
	p-value	0.0613204651	0.0318133533	0.0867269561
	CLEAR result	diff.expr	diff.expr.high.var	diff.expr
	Fold Change	-1.65	-26.94	-2.57
7.2+4.9sox10:GFP array results	*t*-statistic	-6.5065011978	-7.7335410118	-4.5512132645
	p-value	0.0001868201	0.0000556806	0.0018713886
	CLEAR result	diff.expr	diff.expr.high.var	diff.expr
	Fold Change	-1.89	-14.58	-2.53

Table 20: Summary of information for *mitfa*, *dct* and *crestin*.
All tables summarising the information for each validated probe are presented in the same format. Affymetrix gene symbol and GO terms were taken from the NetAffx website. The ZFIN symbol was identified by searching ZFIN with the Affymetrix gene symbol or through a BLAST search using sequence data taken from NetAffx. Orthology to a human gene was determined from either the Ensembl website or the ZFIN website. All array data was taken from the appropriate microarray data set.

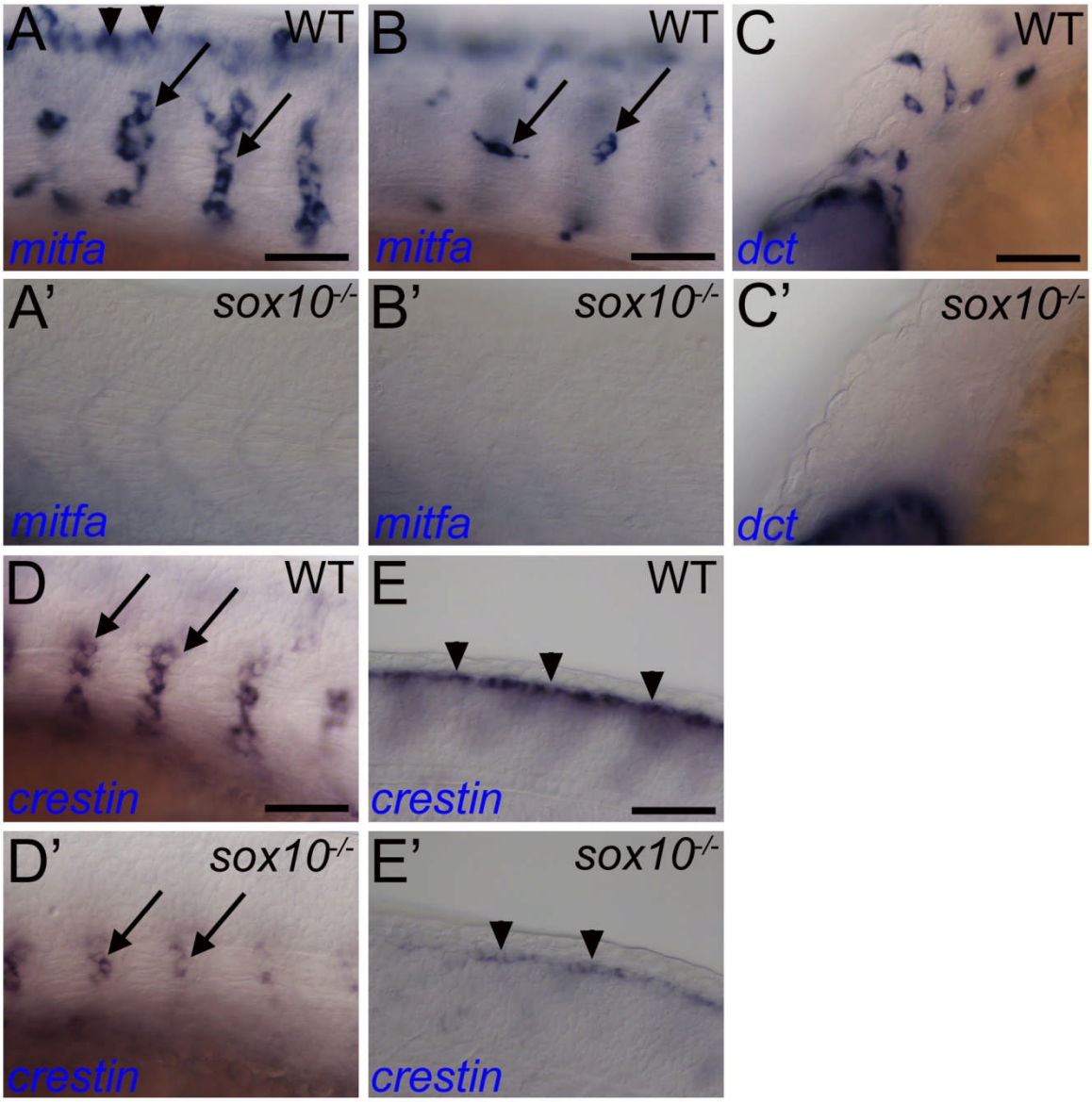


Figure 19: Whole mount *in situ* hybridization expression patterns of *mitfa*, *dct* and *crestin*.
All images are lateral views of 24 hpf zebrafish embryos oriented anterior left and dorsal top. Images labelled with a letter alone are of WT embryos, images labelled with a letter and an apostrophe are corresponding images of *sox10* mutant embryos. Images C and C' are of *dct* expressing melanoblasts in the head, all remaining images are of the trunk. Arrowheads in A, E and E' indicate pre-migratory NCCs. Arrows indicate cells migrating on the medial pathway (A, D and D') and on the lateral pathway (B). Scale bars = 50 µm.

4.2.5.2 Melanocyte Expressed Genes

The following section contains genes that displayed an expression pattern consistent with the melanocyte lineage and were down-regulated in *sox10* mutant embryos. Melanocytes migrate on both the lateral and medial pathways, unlike other NC derivatives that are restricted to migration on a single pathway. Thus gene expression on both pathways is suggestive of expression in the melanocyte lineage. Cells of the melanocyte lineage can also be visualised, using known marker genes, migrating in the midbrain-hindbrain

boundary and over the head. Melanocytes fail to migrate in *sox10* mutant embryos; this is in contrast to neural derivatives as some of these cells still migrate in mutants. The *mitfa* expression pattern typifies a melanocyte expression pattern thus any gene with a similar expression pattern showing the features mentioned above are likely to be expressed in the melanocyte lineage. By 30 hpf, melanoblasts have begun to pigment thus co-localisation of *in situ* hybridization stain with melanin also indicates that the gene is expressed in the melanocyte lineage. While the above characteristics strongly suggest at gene expression in the melanocyte lineage they do not exclude expression in other NCC types; it would require expression of the target gene to be evaluated in zebrafish mutant embryos that specifically lack cells of the melanocyte lineage to confirm this. This would help to distinguish between a melanocyte expressed gene and a gene expressed in two cell types, such as xanthophores and iridophores, that migrate on the lateral and medial pathway respectively.

4.2.5.2.1 Melanocyte expressed V-ATPase genes

Affymetrix Probe Identifier		Dr.2855.1.A1_a_at	Dr.7966.1.S1_at
Affymetrix Gene Symbol		LOC564804 similar to ATPase, H+ transporting, lysosomal 70kDa, V1 subunit A, like	*atp6v1e1*
ZFIN Gene Symbol		*atp6v1a*	*atp6v1e1*
Full Name		*ATPase, H+ transporting, lysosomal V1 subunit A*	*ATPase, H+ transporting, lysosomal, V1 subunit E isoform 1*
Orthology		*ATP6V1A*	*ATP6V1E1*
Biological Process GO Term		N/A	retina development in camera-type eye
Molecular Function GO Term		N/A	hydrogen ion transporting ATPase activity, rotational mechanism
7.2sox10:GFP array results	*t*-statistic	-16.2673625946	-6.6389508247
	p-value	0.0000835642	0.0026715409
	CLEAR result	diff.expr	diff.expr
	Fold Change	-1.86	-1.84
4.9sox10:GFP array results	*t*-statistic	-48.0081	-5.532494545
	p-value	0.0004335995	0.0311520807
	CLEAR result	diff.expr	diff.expr
	Fold Change	-2.33	-2.22
7.2+4.9sox10:GFP array results	*t*-statistic	-13.7012033463	-9.7082853317
	p-value	0.000000776	0.0000105832
	CLEAR result	diff.expr	diff.expr
	Fold Change	-1.96	-1.97

Table 21: Summary of Dr.2855.1.A1_a_at and *atp6v1e1* information.

The Affymetrix probe set Dr.2855.1.A1_a_at corresponds to LOC564804 (similar to ATPase, H+ transporting, lysosomal 70kDa, V1 subunit A, like). This transcript has a 95 % sequence homology with the ZFIN gene *atp6v1a* as identified by a BLAST search on the ZFIN website. This search was performed using the NetAffx sequence data for Dr.2855.1.A1_a_at to interrogate the database of ZFIN RNA and cDNA sequences. *Atp6v1a* was the only hit returned from this BLAST search. It seems likely that probe Dr.2855.1.A1_a_at corresponds to the ZFIN gene *atp6v1a*. This gene has been identified as down-regulated in all three of the microarray data sets by both *t*-Test and CLEAR test (**Table 21**). The Affymetrix probe set Dr.7966.1.S1_at detects transcripts from the *atp6v1e1* gene. This probe set has been identified as down-regulated in all three of the microarray data sets by both the *t*-Test and the CLEAR test (**Table 21**). Thus based on nomenclature, both of these genes encode subunits of the V-ATPase ion transporter.

Cells that express *atp6v1a* were present in a typical pre-migratory NC position along the anterior two thirds of the trunk in 24 hpf WT embryos (**Figure 20A**), expression in this position was reduced in mutant embryos (**Figure 20A'**). Pre-migratory NCCs towards the posterior of the embryo at 30 hpf also expressed *atp6v1a* (data not shown) suggesting that this gene is transiently expressed in early NCCs. In addition *atp6v1a* was expressed in migrating NCCs on both the medial pathway (**Figure 20B** and **E**) and lateral pathway (**Figure 20C** and **D**). In *sox10*$^{-/-}$ embryos no migrating cells expressing *atp6v1a* were detected (**Figure 20B', C', D'** and **E'**). Additionally, some cells expressing *atp6v1a* were also observed to be melanising (**Figure 20D** and **E**). Cranial NCCs expressing *atp6v1a* were also observed in WT embryos in a pattern consistent with a pigment cell fate and again this expression was reduced in mutant embryos (**Figure 20F**, **F'** and data not shown). The expression pattern for *atp6v1e1* (**Figure 21**) was weaker than but very similar to *atp6v1a*. *Atp6v1e1* was expressed in pre-migratory NC at both 24 hpf and 30 hpf stages (**Figure 21A** and **D**), this expression was strongly reduced in *sox10*$^{-/-}$ embryos (**Figure 21A'** and **D'**). *Atp6v1e1* was also expressed in migrating cranial NCCs in the region of the MHB (**Figure 21C**), on the medial pathway (**Figure 21B** and **E**) and on the lateral pathway (**Figure 21F**). Expression in migrating cells particularly in the trunk was weak and hard to visualise. In mutant embryos no expression was detected in migrating cells (**Figure 21B'**, **C'**, **E'** and **F'**) Thus it seems likely that both genes are expressed in melanoblasts and pigmenting melanocytes. Mucus secreting cells also expressed both of these V-ATPase genes, these cells can be distinguished from NCCs by their more ventral position, more rounded (not dendritic) morphology and qualitatively higher levels of V-ATPase expression (data not shown). Mucus secreting cells did not display any change in expression level in *sox10*$^{-/-}$ embryos (data not shown).

There is no embryonic expression pattern for *atp6v1a* or *atp6v1e1* published on the ZFIN website for comparison. An insertion mutant (hi577a) has generated a lesion in

atp6ve1e, this mutant has a phenotype of severely reduced body pigment at 2 dpf and by 3 dpf abnormal spotty pigment, a small head and small otoliths (Amsterdam et al., 2004). As this gene is likely to be expressed in melanoblasts and melanocytes, this mutant pigment phenotype fits with the observed pattern. In summary, *atp6v1a* and *atp6v1e1* are both likely to be expressed in the melanocyte lineage and are both down-regulated in *sox10* mutant embryos. *Atp6v1a* and *atp6v1e1* both represent novel NC expressed genes.

The protein product of *atp6v1a* and *atp6v1e1* are subunits of the V_1 sector of the vacuolar-H^+ ATPase (V-ATPase). V-ATPase is a complex multi-subunit enzyme that couples proton pumping to ATP hydrolysis; the V_0 sector is responsible for proton translocation across membranes while the V_1 sector contains the catalytic sites necessary for ATP hydrolysis (Pietrement et al., 2006, Tabata et al., 2008). V-ATPases are responsible for the acidification of intracellular vesicles such as lysosomes. With *atp6v1a* and *atp6v1e1* expressed in melanoblasts and melanocytes, this protein complex may play a role in the acidification of melanosomes. With two genes that encode subunits of the same proton pump validated during the *in situ* hybridization screen, the microarray data was examined for further examples of V-ATPase genes down-regulated in *sox10*$^{-/-}$ embryos (**Table 22**). A further seven V-ATPase genes were indicated as differentially regulated in one or more of the microarray data sets by *t*-Test or CLEAR test. Thus a total of nine V-ATPase genes have been identified as differentially regulated in *sox10*$^{-/-}$ embryos. As all these genes are subunits in the same proton pump, it is the function of this proton pump in melanocyte development that is of interest. As stated previously, the V-ATPase proton pump might play a role in regulating the acidity of melanosomes. The prevalence of V-ATPase genes in the microarray data and their possible function in the NC is examined further in the discussion.

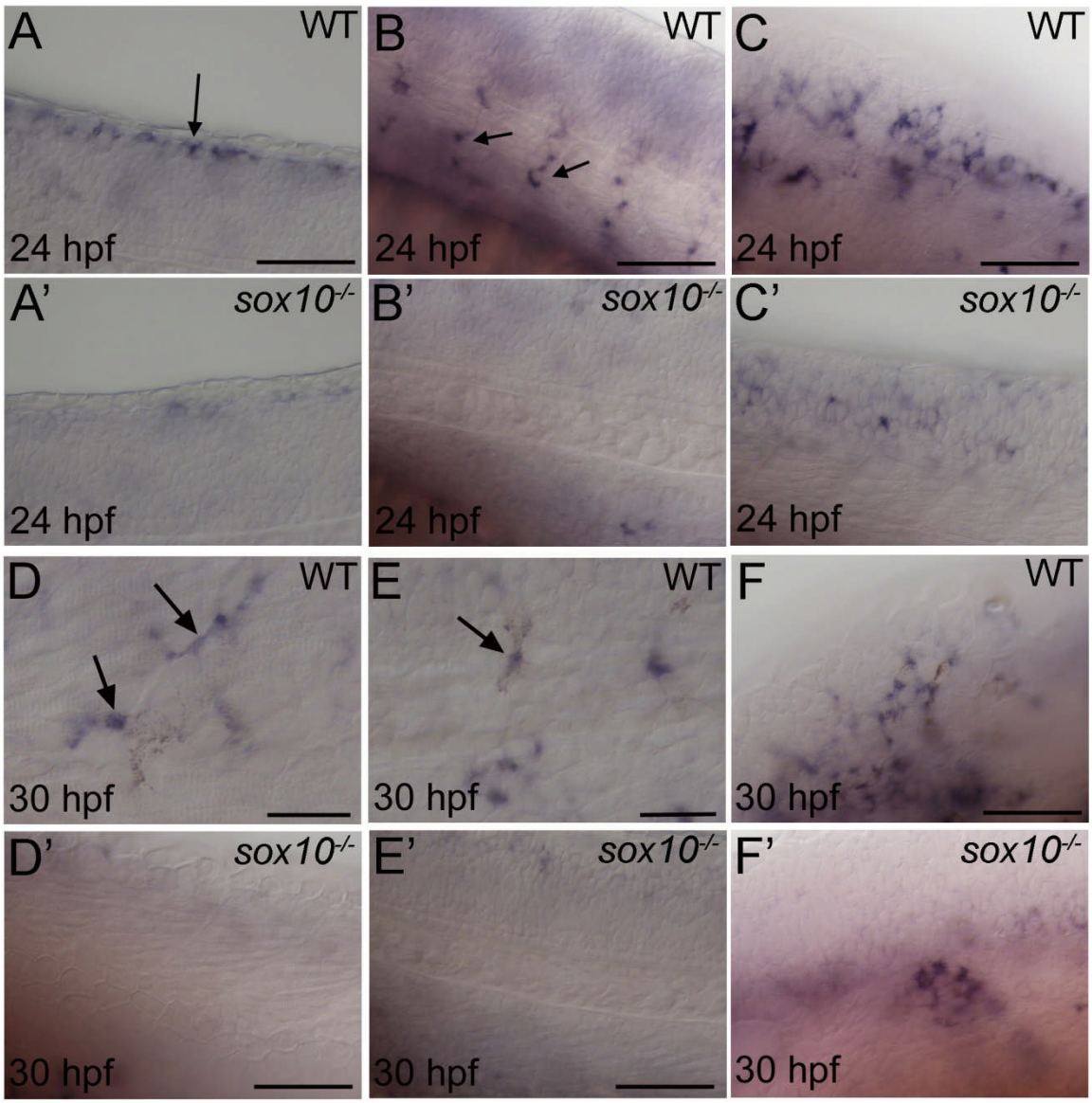


Figure 20: Whole mount *in situ* hybridization pattern of Dr.2855.1.A1_a_at expression.

All images are lateral views of embryos oriented anterior to the left, dorsal top, posterior to the right and ventral bottom. Images A-C' are of 24 hpf embryos and images D-F' are of 30 hpf embryos, A-F are of WT embryos, A'-F' are corresponding images of *sox10*$^{-/-}$ embryos. A and A' are images of dorsal mid-trunk, (A) arrow indicates pre-migratory NC expressing *atp6v1a*. Expression of *atp6v1a* is strongly reduced in mutant embryos. Images B, B', C and C' are of anterior trunk, in B and B' focus is on the interior of the embryo, cells migrating on the medial pathway (arrows in B) express *atp6v1a*. (C) Cells expressing *atp6v1a* are visible just underneath the epidermis and are migrating on the lateral pathway. *Atp6v1a* expression cannot be detected in migrating NCCs in mutant embryos. Images D, D', E and E' are of mid-trunk, (D) arrow indicates a cell migrating on the lateral pathway expressing *atp6v1a* co-localised with melanin. (E) Arrow points to a cell on the medial pathway expressing *atp6v1a* co-localised with melanin. Images F and F' are of the posterior head immediately behind the otic vesicle, cells in F expressing *atp6v1a* are becoming pigmented with melanin. Scale bars in D and E = 25 µm, all other scale bars = 50 µm.

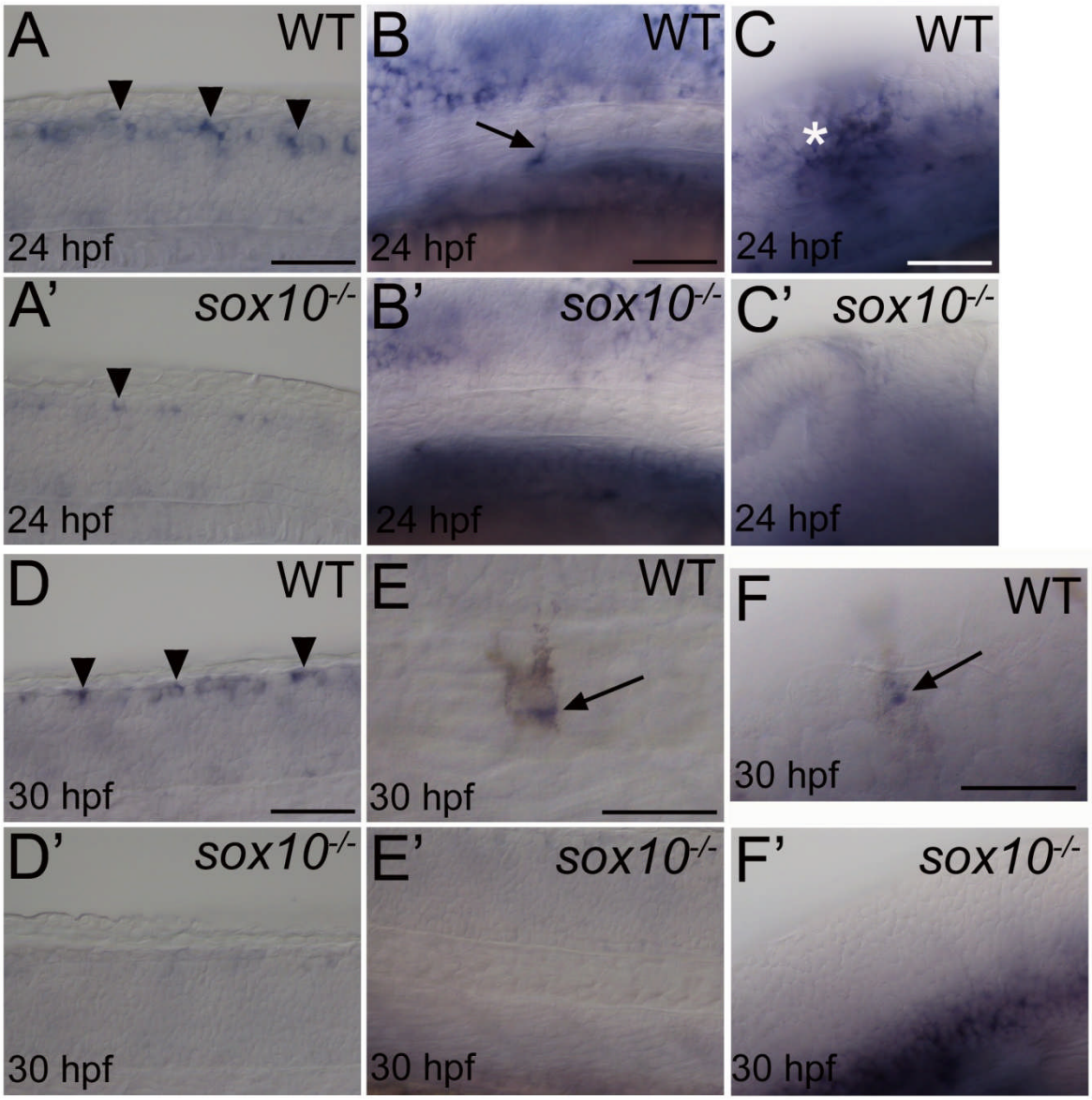


Figure 21: Whole mount *in situ* hybridization pattern of *atp6v1e1* expression.
All images are presented as per Figure 20. Images C and C' are of the head at the midbrain-hindbrain boundary (MHB), images F and F' are of posterior head and anterior trunk, all other images are of trunk. Images A and D show expression of *atp6v1e1* in a pre-migratory NC position (arrowheads), images A' and D' are of mutant embryos in the same position showing reduced levels of expression. Image C shows expression in NCCs in the region of the MHB (*), expression was reduced in mutant embryos (C'). Images B and E show cells expressing *atp6v1e1* on the medial pathway (arrows), in E expression co-localises with melanin. Images B' and E' show that no cells expressing *atp6v1e1* were migrating on the medial pathway in mutant embryos. Image F shows a cell expressing *atp6v1e1* co-localised with melanin (arrow) migrating just under the epidermis. Scale bars in E and F = 25 µm, all other scale bars = 50 µm.

Affymetrix Identifier	Gene Symbol	7.2 Array Data		4.9 Array data		7.2+4.9 Array Data	
		p-value	CLEAR	*p*-value	CLEAR	*p*-value	CLEAR
Dr.14783.1.S1_at	*atp6v1c1*	0.044	diff.expr	0.020	diff.expr	0.032	diff.expr
Dr.186.1.S1_at	*atp6v1ba*	0.042	diff.expr	0.009	non.sig	0.001	diff.expr
Dr.14072.1.S1_at	*atp6v1d*	0.011	diff.expr	0.136	non.sig	0.002	diff.expr
Dr.1971.1.A1_at	*atp6v0c*	0.009	diff.expr	0.136	non.sig	0.0003	non.sig
Dr.370.1.S1_at	*atp6v1g1*	0.005	diff.expr	0.005	diff.expr	0.00005	diff.expr
Dr.13658.1.S1_at	*atp6v1f*	0.063	diff.expr	0.035	diff.expr	0.001	diff.expr
Dr.14847.1.A1_at	*atp6v1c1l*	0.066	diff.expr.high.var	0.135	diff.expr	0.004	diff.expr

Table 22: Summary of Differentially Expressed V-ATPase genes.
Genes named as V-ATPases by Affymetrix that were identified as differentially regulated in one or more microarray data sets are summarised. Results from all three microarray data sets are given for the t-Test (p-value) and CLEAR test. Statistically significant results (differentially regulated) are highlighted in green; results that are not significant are highlighted in red**.**

4.2.5.2.2 Dr.8594.1.S1_at

Affymetrix Probe Identifier		Dr.8594.1.S1_at
Affymetrix Gene Symbol		*zgc:100919*
ZFIN Gene Symbol		*zgc:100919*
Full Name		*zgc:100919*
Orthology		N/A
Biological Process GO Term		transport
Molecular Function GO Term		transporter activity, lipid binding
7.2sox10:GFP array results	*t*-statistic	-7.944715023
	p-value	0.0013592674
	CLEAR result	diff.expr
	Fold Change	-1.81
4.9sox10:GFP array results	*t*-statistic	-4.4917368889
	p-value	0.0461597666
	CLEAR result	diff.expr
	Fold Change	-2.00
7.2+4.9sox10:GFP array results	*t*-statistic	-11.8869142532
	p-value	0.0000023042
	CLEAR result	diff.expr
	Fold Change	-1.89

Table 23: Summary of Dr.8594.1.S1_at information

The Affymetrix probe set Dr.8594.1.S1_at corresponds to the gene *zgc:100919.* Currently this gene is only known by a Zebrafish Gene Collection code and has yet to be identified with a gene name. According to the NetAffx website and the Entrez Gene entry for *zgc:100919*, this gene may be a member of the tetraspanin superfamily and specifically has similarity to the subfamily of vertebrate transmembrane 4 superfamily 8 (TM4SF8) like genes. Dr.8594.1.S1_at was consistently detected as down-regulated in all three microarray data sets (**Table 23**). This gene is expressed in a sub-set of pre-migratory

NCCs along most of the trunk at 24 hpf (**Figure 22A**). By 30 hpf pre-migratory expression was restricted to the posterior trunk only (data not shown). Cells migrating on the medial pathway at 24 hpf (**Figure 22A**) and at 30 hpf (data not shown) also expressed *zgc:100919*. Cells expressing *zgc:100919* were also detected migrating on the lateral pathway at both time points (data not shown). Expression in pre-migratory and migrating cells was absent in *sox10*$^{-/-}$ embryos (**Figure 22A'** and **B'**). At 30 hpf *zgc:100919* expression overlapped with cells that were beginning to pigment with melanin (**Figure 22B** and **C**). *Zgc:100919* thus appeared to be expressed in early NCCs, specifically, melanoblasts and melanocytes and was down-regulated in *sox10* mutant embryos. An expression pattern for *zgc:100919* has been published on the ZFIN website and this shows expression in melanoblasts at 24 hpf with expression continuing in melanocytes, at least until 48 hpf (Thisse et al., 2004). The probe used to detect *zgc:100919* mRNA in this project has generated a strong signal with very little background expression. The expression pattern generated appears to mark a large number of melanoblasts, a similar number as marked by a *mitfa in situ* hybridization probe. These two facts make this an attractive marker gene for the melanocyte lineage.

Tetraspanins are a family of transmembrane proteins that have been associated with the regulation of a number of cellular processes including cell proliferation, motility and cell death (Hemler, 2005). These are all processes important during NC development but despite this tetraspanins have not been strongly linked to NC biology. A study has identified that a number of tetraspanins localise to the end of projections that extend from cultured melanocytes. This includes at the end of projections that connect to other melanocytes and at the tip of dendrites that act during cell migration (García-López et al., 2005). Knockdown of tetraspanins reduced the ability of these melanocytes to migrate. It would be of interest to assess the knockdown of *zgc:100919* in zebrafish embryos, perhaps focusing on defects in melanocyte migration. Such a study would be particularly important given the context that very little study of tetraspanins in zebrafish has been undertaken (Hemler, 2005).

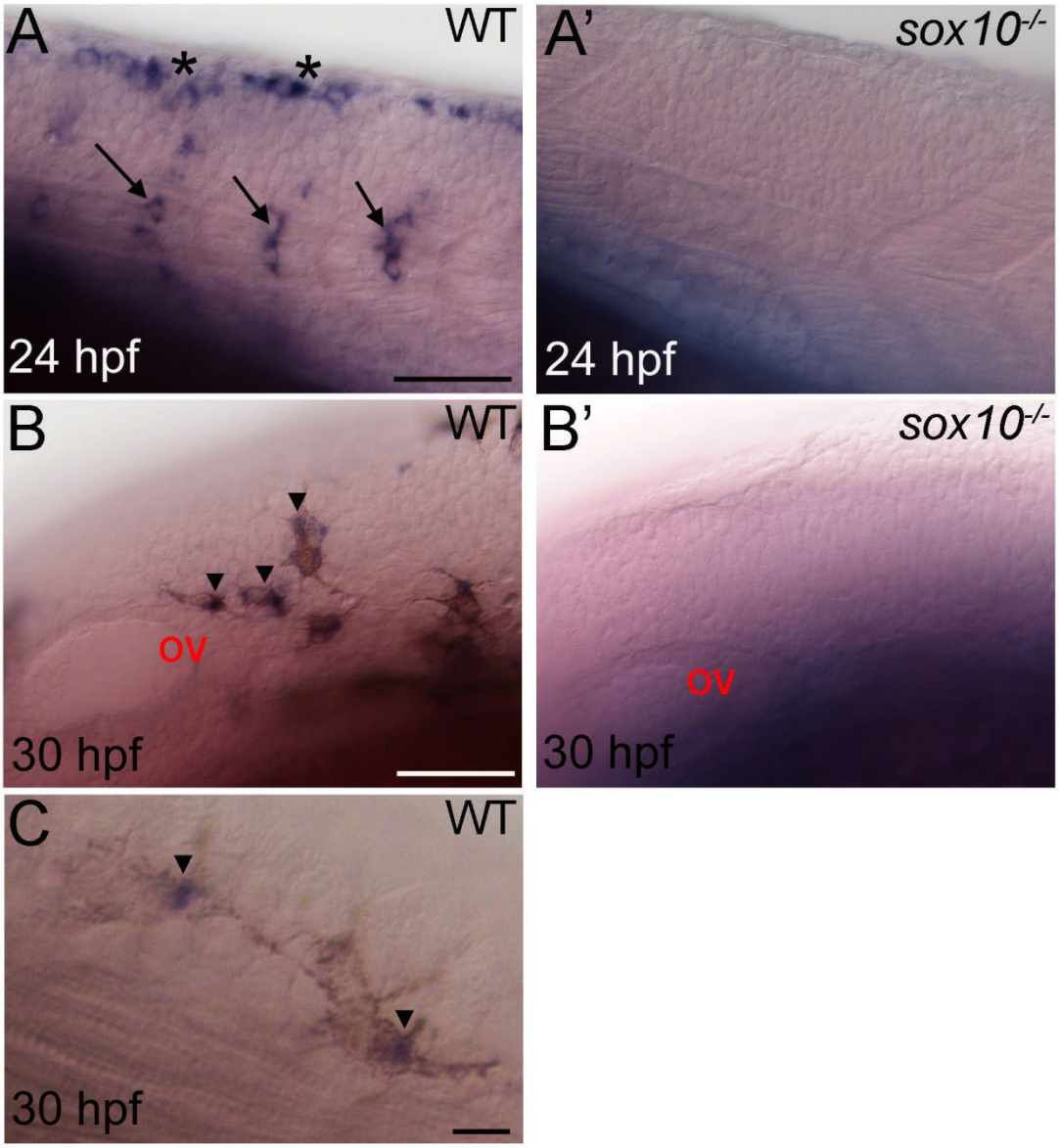


Figure 22: Whole mount *in situ* hybridization expression pattern of *zgc:100919*.
All images presented as per Figure 20. Images A and A' are of the trunk of 24 hpf embryos while B, B' and C are images of 30 hpf embryos. B and B' are images of the posterior head just behind the otic vesicle (ov). C is a magnified image of two melanised cells in the trunk. *Zgc:100919* was expressed in a subset of pre-migratory NCCs (asterisks in A) and cells migrating on the medial pathway (arrows in A). At 30 hpf *zgc:100919* expression overlapped with melanising cells (arrowheads in B and C). Expression of *zgc:100919* was strongly reduced in *sox10* mutant embryos (A' and B'). Scale bars in A and B = 50 µm, scale bar in C = 10 µm.

4.2.5.2.3 *pah*

Affymetrix Probe Identifier		Dr.4171.1.A1_at
Affymetrix Gene Symbol		*pah*
ZFIN Gene Symbol		*pah*
Full Name		*phenylalanine hydroxylase*
Orthology		*PAH*
Biological Process GO Term		L-phenylalanine catabolic process, metabolic process, aromatic amino acid family metabolic process
Molecular Function GO Term		monooxygenase activity phenylalanine 4-monooxygenase activity, iron ion binding, amino acid binding
7.2sox10:GFP array results	*t*-statistic	-7.4460177422
	p-value	0.0017376344
	CLEAR result	diff.expr
	Fold Change	-3.70
4.9sox10:GFP array results	*t*-statistic	-14.469004631
	p-value	0.0047426876
	CLEAR result	diff.expr
	Fold Change	-3.07
7.2+4.9sox10:GFP array results	*t*-statistic	-8.8098163605
	p-value	0.0000216759
	CLEAR result	diff.expr
	Fold Change	-3.30

Table 24: Summary of *pah* Information

The Affymetrix probe set Dr.4171.1.A1_at has been annotated as *pah*, analysis of all three microarray data sets has shown that this gene is down-regulated (**Table 24**). Analysis by *in situ* hybridization showed that *pah* is expressed in pre-migratory NCCs (**Figure 23B**). A *sox10 in situ* hybridization at the same stage would label more NCCs therefore only a subset of pre-migratory NCCs is labelled. Migrating cells on both the medial pathway (**Figure 23B** and **D**) and on the lateral pathway (**Figure 23C**) also express *pah*. Expression in all NCCs was strongly reduced in *sox10*$^{-/-}$ embryos although some pre-migratory cells still expressed *pah* (**Figure 23B'** and **C'**). At 30 hpf *pah* expression overlapped with lightly melanised melanocytes (**Figure 23E** and **F**). Thus *pah* appears to be expressed in both melanoblasts and melanocytes. Interestingly, in contrast with most other melanocyte expressed genes, *pah* expression is still detectable in some pre-migratory mutant NCCs (**Figure 23B'** and **C'**). This is a unique result among the melanocyte expressed genes assessed during this screen. It might be that *pah* expression does not solely rely on Sox10 induction of *mitfa* expression. Alternatively, evidence in zebrafish suggests that Sox10 can repress expression of some melanocyte markers that are down stream of *mitfa* (Greenhill, 2008). Continued expression of *pah* in *sox10* mutant embryos may result from a loss of Sox10 depression although expression levels are actually reduced as *mitfa* expression is not induced. The level of *pah* expression seen in

sox10 mutant embryos is much higher than seen with residual *dct* expression (**Figure 19**). *Pah* has been annotated as expressed in pre-migratory and migrating NCCs and pigment cells (Thisse et al., 2004) in a pattern that suggests this gene is a marker of the melanocyte lineage. *Pah* codes for an enzyme, phenylalanine hydroxylase, which plays an important role in melanogenesis. PAH catalyses the conversion of L-phenylalanine to L-tyrosine which is the substrate for the key melanogenic enzyme tyrosinase (Schallreuter et al., 2008). The expression pattern of *pah* observed was to be expected based on previous *in situ* hybridization patterns and the role of PAH in melanin synthesis. In summary, *pah* was expressed in cells of the melanocyte lineage of WT embryos and strongly down-regulated in *sox10* mutant embryos thus validating this microarray result.

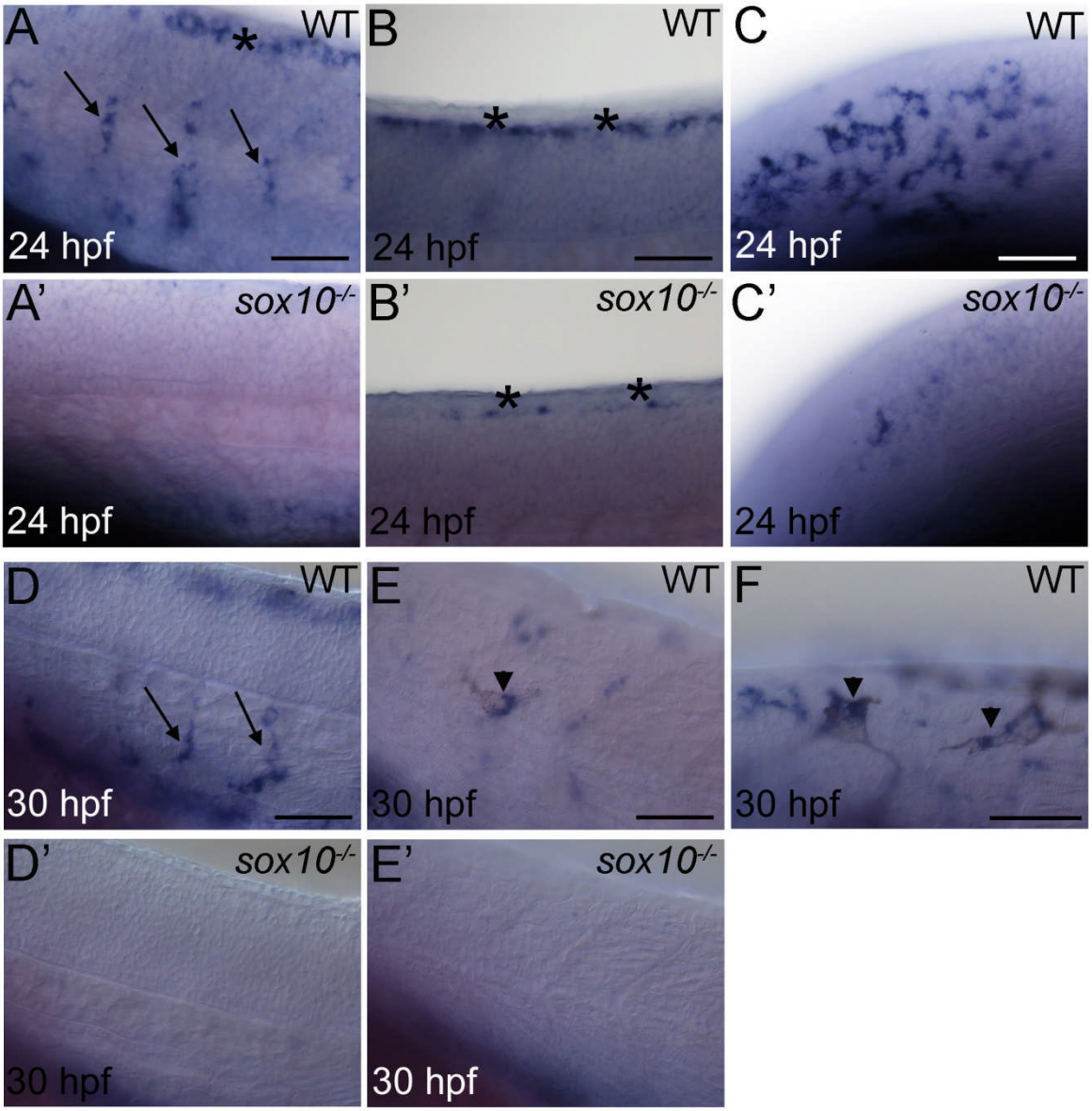


Figure 23: Whole mount *in situ* hybridization pattern of *pah* expression.
Figure is presented as per Figure 20. A, A' and D-E' show anterior trunk with pre-migratory NCCs indicated by asterisks (*), cells migrating on the medial pathway are shown using arrows and a lightly melanised cell on the lateral pathway highlighted with an arrowhead. Images B and B' are of mid-trunk, pre-migratory NCCs are indicated with an asterisk. Images C and C' show posterior head and anterior trunk with NCCs migrating on the lateral pathway. F shows lightly melanised cells co-expressing *pah* (arrowheads) in the region of the most anterior somites. Scale bar in F = 25 µm, scale bars in all other images = 50 µm.

4.2.5.2.4 *degs1*

Affymetrix Probe Identifier		Dr.1251.1.S1_at
Affymetrix Gene Symbol		*degs1*
ZFIN Gene Symbol		*degs1*
Full Name		degenerative spermatocyte homolog, lipid desaturase
Orthology		*DEGS1*
Biological Process GO Term		lipid metabolic process oxidation reduction
Molecular Function GO Term		oxidoreductase activity
7.2sox10:GFP array results	*t*-statistic	-6.2696032524
	p-value	0.0033029604
	CLEAR result	diff.expr
	Fold Change	-1.58
4.9sox10:GFP array results	*t*-statistic	-2.097016573
	p-value	0.1709180772
	CLEAR result	diff.expr
	Fold Change	-1.50
7.2+4.9sox10:GFP array results	*t*-statistic	-2.9583206177
	p-value	0.0181928389
	CLEAR result	diff.expr
	Fold Change	-1.54

Table 25: Summary of *degs1* information

The Affymetrix probe Dr.1251.1.S1_at corresponds to the zebrafish gene *degs1*. The CLEAR test identified *degs1* as differentially regulated in all three data sets while the *t*-Test identified differential regulation in two data sets but not the 4.9 data set (**Table 25**). Whole mount *in situ* hybridization identified *degs1* expression in cells migrating on the medial pathway (**Figure 24A** and **C**) and lateral pathway (**Figure 24B**). Expression in these migrating cells was absent in *sox10*$^{-/-}$ embryos (**Figure 24A'**, **B'** and **C'**). Weak *degs1* expression was also detected in a few pre-migratory NCCs which also displayed reduced levels of *degs1* expression in mutant embryos (data not shown). *Degs1* expression co-localised with cells that were pigmenting with melanin (**Figure 24C**). In general, *degs1* expression was weak but it appears to be expressed in melanoblasts and pigmenting melanocytes at least until 30 hpf. An expression pattern for *degs1* displaying expression in migrating NCCs which shows strong similarities to the pattern shown in **Figure 24** has been submitted to the ZFIN website (Thisse et al., 2004). This expression pattern also identified *degs1* expression in the otic vesicle at around 24 hpf. Otic vesicle expression was not observed during the current study but this could have been a result of weak staining. In conclusion, d*egs1* appears to be expressed in cells of the melanocyte lineage and is down-regulated in *sox10* mutant embryos.

Very little literature is available regarding *degs1* however the gene was first cloned and named as *MLD*, a membrane fatty acid (lipid) desaturase (Cadena et al., 1997). When transfecting a human embryonic kidney cell line with *MLD* cDNA the authors noted

that epidermal growth factor (EGF) receptor expression was inhibited by MLD. A study of the *in vitro* effects of *DEGS1* expression in human cancer cell lines identified a role in progression of the cell cycle from G1 to S phase. This role was dependent on NF-κB activation of cyclin-D1 expression (Zhou et al., 2009). Over expression of DEGS1 also promoted the migration of Eca109 oesophageal carcinoma cells (Zhou et al., 2009). Thus while a role for *degs1* in neural crest biology has yet to be established, this gene is known to play a role in cell proliferation and migration, two important processes that NCCs are actively involved in.

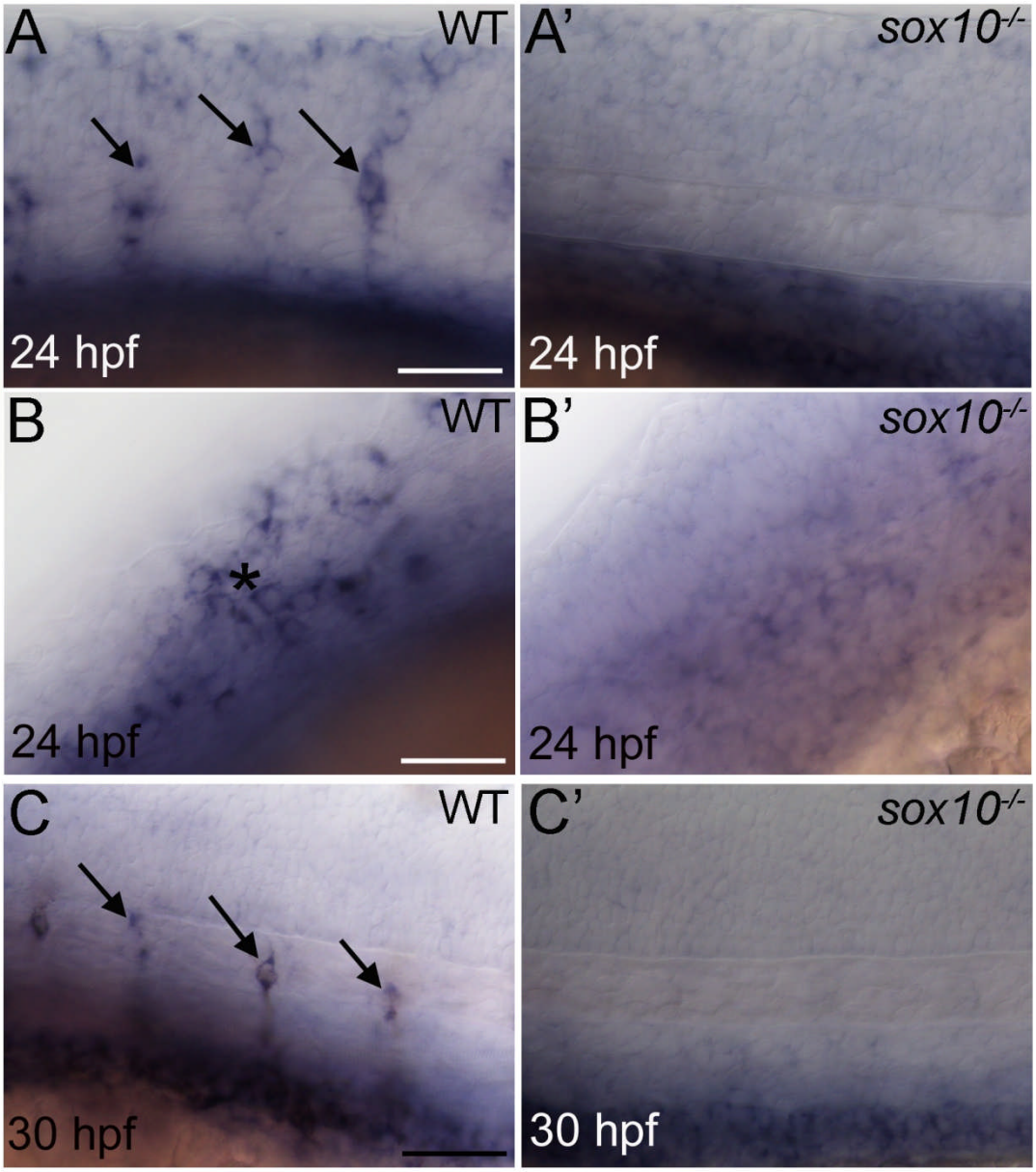


Figure 24: Whole mount *in situ* hybridization pattern of *degs1* expression.
Images are presented as per Figure 20. *Degs1* expression was detected in cells migrating on the medial pathway in anterior trunk (arrows in A and C), this expression was absent in mutant embryos (A' and C'). Expression was also detected in WT embryos on the lateral pathway in cells migrating at the posterior head/anterior trunk axial level (* in B), this expression was absent in *sox10*$^{-/-}$ embryos (B'). Expression of *degs1* co-localised with cells synthesising melanin (C). Scale bars = 50 µm.

4.2.5.2.5 *zgc:110239*

Affymetrix Probe Identifier		Dr.7250.1.A1_at
Affymetrix Gene Symbol		*zgc:110239*
ZFIN Gene Symbol		*zgc:110239*
Full Name		*zgc:110239*
Orthology		N/A
Biological Process GO Term		proteolysis
Molecular Function GO Term		cysteine-type endopeptidase activity cysteine-type peptidase activity
7.2sox10:GFP array results	*t*-statistic	-6.6761045456
	p-value	0.002616606
	CLEAR result	diff.expr
	Fold Change	-1.41
4.9sox10:GFP array results	*t*-statistic	-0.569756031
	p-value	0.626308918
	CLEAR result	non.significant
	Fold Change	-1.09
7.2+4.9sox10:GFP array results	*t*-statistic	-1.1016930342
	p-value	0.302633673
	CLEAR result	non.significant
	Fold Change	-1.21

Table 26: Summary of Information for *zgc:110239*

The Affymetrix probe set Dr.7250.1.A1_at corresponds to the gene *zgc:110239*. Both statistical tests identified this gene as differentially regulated in the 7.2 data set. However this gene was not identified as differentially regulated in the 4.9 or the 7.2+4.9 data sets (**Table 26**). To validate if *zgc:110239* was down-regulated, *in situ* hybridization was performed (**Figure 25**). In 24 hpf WT embryos *zgc:110239* was expressed in positions typical of cranial NCCs at the midbrain-hindbrain boundary (data not shown) and near the border between posterior head and anterior trunk (**Figure 25A**). *Zgc:110239* expression was also observed in a few pre-migratory NCCs in the trunk (**Figure 25B**). Expression was absent from these sites in *sox10*$^{-/-}$ embryos (**Figure 25A'**, **B'** and data not shown). In 30 hpf embryos, *zgc:110239* was expressed in pre-migratory NC positions in WT embryos, this expression was reduced but not absent in mutant embryos (**Figure 25D** and **D'**). *Zgc:110239* expression also overlapped with melanising cells migrating in WT embryos (**Figure 25C** and **E**), this expression was absent in *sox10* mutant embryos (**Figure 25C'** and **E'**). The expression pattern examined here was weak and therefore hard to interpret but clear NCC expression was observed that was reduced in mutant embryos. The expression pattern of *zgc:110239* has also been determined and published on the ZFIN website (Thisse et al., 2004). This pattern shows extensive *zgc:110239* expression in pre-migratory NCCs and migrating NCCs. While not explicitly annotated as such, it appears that the cells are migrating on both the medial and lateral pathways. The *in situ* hybridization pattern from Thisse et al., 2004 is stronger and clearer than the pattern shown in **Figure 25**. There is currently no work that hints at a possible functional

role for this gene in the NC. In summary, *zgc:110239* is expressed in NCCs of the melanocyte lineage and is down-regulated in *sox10* mutant embryos.

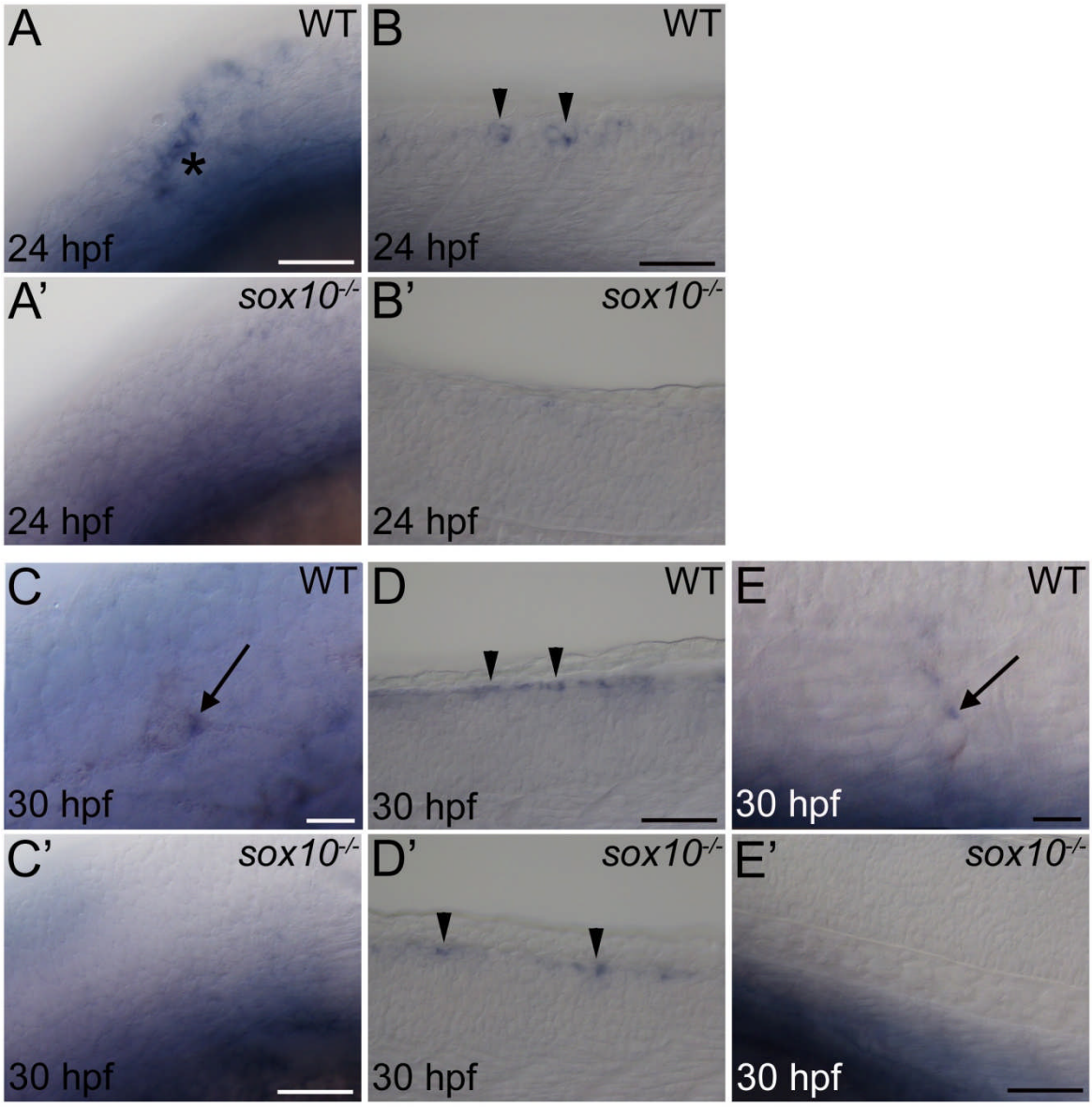


Figure 25: Whole mount *in situ* hybridization pattern of *zgc:110239* expression.
All images are lateral views of zebrafish embryos presented as per Figure 20. Images A, A', C and C' are of posterior head and anterior trunk; all other images are of mid-trunk. Expression of *zgc:110239* in 24 hpf WT embryos was observed in a NC like position in the head (asterisk in A) and pre-migratory NCCs in the trunk (arrowheads in B). Expression was absent from these sites in mutant embryos (A' and B'). At 30 hpf, *zgc:110239* was expressed in pre-migratory NCCs (D), this expression was seen in *sox10* mutant embryos (D'). *Zgc:110239* expression also overlapped with migrating melanising cells at 30 hpf in WT embryos (arrows in C and E). No migrating cells expressed *zgc:110239* in mutant embryos (C' and E'). Scale bars in C and E = 10 µm, all other scale bars = 50 µm.

4.2.5.3 Xanthophore Expressed Genes

Xanthophores are yellow pigment cells which become pigmented from around 42 hpf, considerably later than melanocytes which are visibly pigmented from 28 hpf. Despite this, expression of early xanthophore lineage markers such as *gch* and *xdh* can be visualised at 24 hpf, when early melanocyte genes such as *mitfa* and *dct* are expressed in melanoblasts. Xanthoblasts can be visualised in pre-migratory NC positions and then migrating on the lateral pathway only. In addition xanthoblasts can be detected migrating over the head of zebrafish embryos. This cannot exclude that a gene might be expressed in a sub-set of melanocytes that are restricted to migrating on the lateral pathway, although there is currently no precedent for such a gene. In order to examine this, a xanthophore marker should not show extensive overlap with melanising cells at 30 hpf. Expression of early xanthophore markers is strongly reduced but still detectable in *sox10*$^{-/-}$ embryos. In particular, xanthoblast markers are expressed in a reduced number of NCCs stuck in clusters in pre-migratory positions in the trunk and head. The above criteria, particularly expression of the target gene in cells migrating on the lateral pathway only are indicative of the xanthophore lineage and will be used during this study.

4.2.5.3.1 *rbp4l*

Affymetrix Probe Identifier		Dr.10292.1.S1_at
Affymetrix Gene Symbol		*rbp4l*
ZFIN Gene Symbol		*rbp4l* (*purpurin*)
Full Name		*retinol binding protein 4, like*
Orthology		None available
Biological Process GO Term		Transport
Molecular Function GO Term		transporter activity, retinoid binding
7.2sox10:GFP array results	*t*-statistic	-16.037355423
	p-value	0.0000883988
	CLEAR result	diff.expr
	Fold Change	-7.77
4.9sox10:GFP array results	*t*-statistic	-41.8679580688
	p-value	0.0005699871
	CLEAR result	diff.expr
	Fold Change	-18.65
7.2+4.9sox10:GFP array results	*t*-statistic	-12.7175235748
	p-value	0.0000013756
	CLEAR result	diff.expr.high.var
	Fold Change	-11.03

Table 27: Summary of *rbp4l* information.

The Affymetrix probe Dr.10292.1.S1_at detects transcripts from the *rbp4l* gene, which has been known previously as *purpurin* (Tanaka et al., 2007). This transcript was detected as down-regulated in all three microarray data sets by both statistical tests (**Table 27**). The

fold change values for this transcript were some of the highest detected suggesting a very strong reduction in expression. Surprisingly then, whole mount *in situ* hybridization did not show any detectable expression at 24 hpf but strong expression was detected at 30 hpf (**Figure 26**). With strong down-regulation of this gene recorded by microarray analysis this suggested that the FACS enriched cells were slightly older than 24 hpf, from a time point when *rbp4l* was highly expressed. *Rbp4l* expression was detected in cells migrating on the lateral pathway (**Figure 26A** and **E**) but not on the medial pathway (**Figure 26C**). Expression was also detected in a sub-set of pre-migratory NCCs (**Figure 26C**) and some cranial NCCs (data not shown). No NCC expression was detected in *sox10*$^{-/-}$ embryos (**Figure 26B** and **D**). *Rbp4l* was also expressed strongly in the epiphysis of WT and mutant embryos equally at both time points (data not shown). The expression pattern for *rbp4l* has already been described (Thisse et al., 2004) and shows strong similarities to the pattern presented here. The data suggests that *rbp4l* is expressed in the xanthophore lineage prior to these cells becoming visibly pigmented with detectable expression first occurring at a time point between 24 and 30 hpf. *Rbp4l* expression is strongly reduced in *sox10* mutant embryos in comparison to WT embryos thus validating this microarray result.

The zebrafish *rbp4l* gene has been identified as an orthologue of the chick and goldfish *purpurin* gene (Tanaka et al., 2007). Purpurin is a secreted molecule that has been shown to promote both cell adhesion and survival of neural retinal cells in culture (Schubert and LaCorbiere, 1985, Schubert et al., 1986). *Purpurin* is expressed in the developing retina of zebrafish and also in adult fish photoreceptor cells during regeneration after optic nerve transection (Matsukawa et al., 2004, Tanaka et al., 2007). Neurite outgrowth from retinal explants was promoted by simultaneous addition of retinol and Purpurin (Matsukawa et al., 2004). Thus while the role of *purpurin* or *rbp4l* is now becoming understood in adult retinas, no biological function has been attributed to this gene during the development of the NC. As *rbp4l* is expressed later than early xanthoblast markers this suggests that *rbp4l* does not play a role in xanthoblast specification. As *rbp4l* is strongly expressed in migrating cells this gene may play a role in cell migration, possibly by mediating cell to cell adhesion. Alternatively *rbp4l* may promote xanthoblast survival. A morpholino mediated knock down of *rpb4l* expression may help to elucidate a function for this gene in the NC.

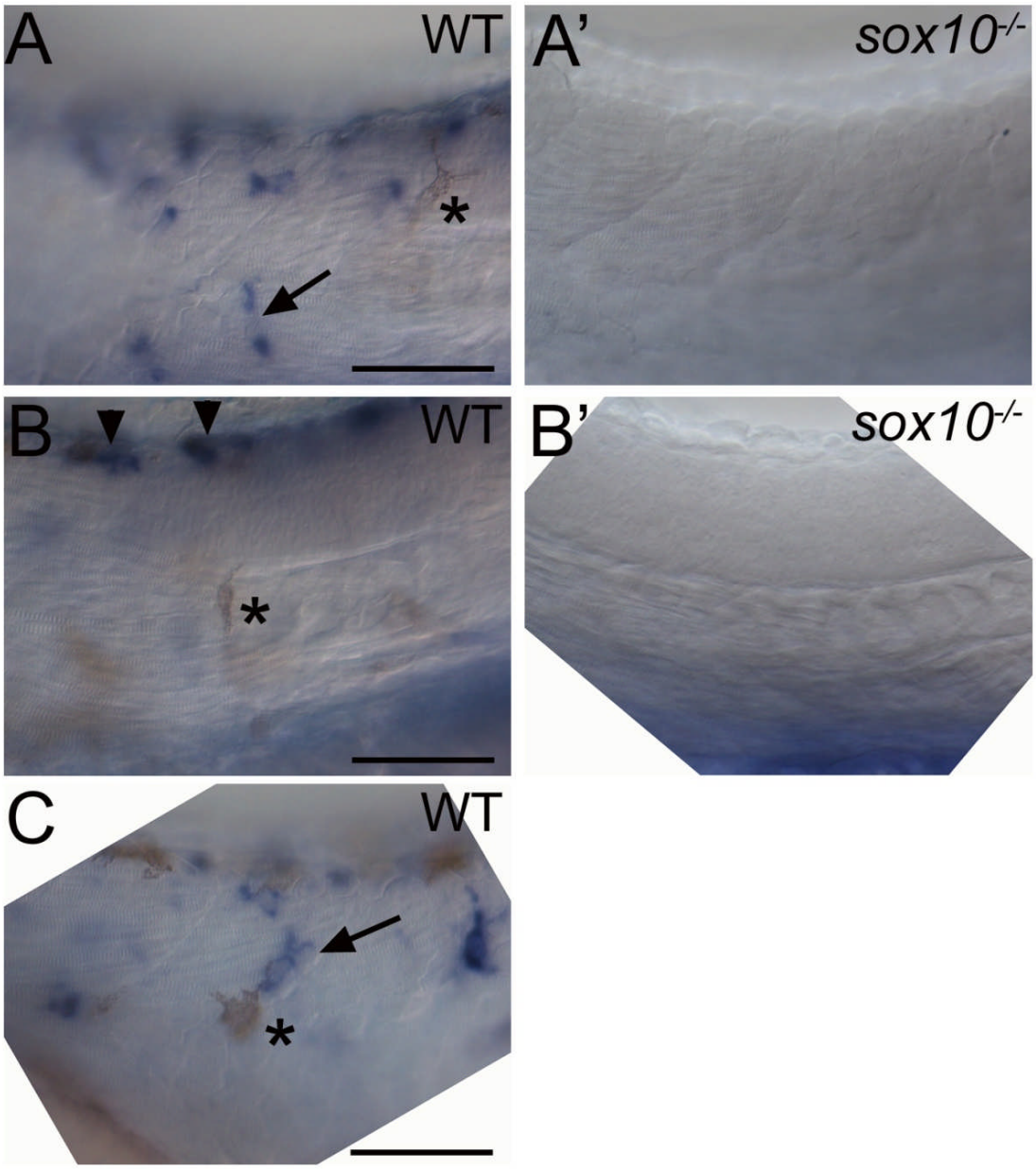


Figure 26: Whole mount *in situ* hybridization pattern of *rbp4l* expression.
Images are presented as per Figure 20. All images are of 30 hpf embryos in the mid-trunk region. Cells expressing *rbp4l* were visible migrating on the lateral pathway (arrows in A and C) in WT embryos but no expression was detected in mutant embryos (A'). *Rbp4l* was also expressed in WT pre-migratory NCCs (arrowheads in B) but absent in mutant embryos (A' and B'). No cells expressing *rbp4l* were detected migrating on the medial pathway (B) and no melanising cells expressed *rbp4l* (asterisks in A, B and C). Scale bars = 50 µm.

4.2.5.3.2 *pax7*

Affymetrix Probe Identifier		Dr.23456.1.S1_at
Affymetrix Gene Symbol		*pax7*
ZFIN Gene Symbol		*pax7a*
Full Name		*paired box gene 7a*
Orthology		*PAX7*
Biological Process GO Term		transcription transcription termination regulation of transcription, DNA-dependent multicellular organismal development transcription antitermination regulation of transcription
Molecular Function GO Term		DNA binding transcription factor activity sequence-specific DNA binding
7.2sox10:GFP array results	*t*-statistic	-10.5922355652
	p-value	0.0004495994
	CLEAR result	non.significant
	Fold Change	-1.25
4.9sox10:GFP array results	*t*-statistic	-0.4997960031
	p-value	0.6667875648
	CLEAR result	non.significant
	Fold Change	-1.02
7.2+4.9sox10:GFP array results	*t*-statistic	-4.6060390472
	p-value	0.001741653
	CLEAR result	non.significant
	Fold Change	-1.22

Table 28: Summary of *pax7* information.

The Affymetrix probe set Dr.23456.1.S1_at corresponds to the transcription factor *pax7*. This gene was identified as differentially regulated in the 7.2 and 7.2+4.9 data sets by the *t*-Test, this gene was not identified as differentially regulated by the CLEAR test (**Table 28**). *Pax7* was expressed in pre-migratory NCCs dorsal of the neural tube in 24 hpf and 30 hpf WT embryos and *sox10*$^{-/-}$ embryos (**Figure 27A**, **A'** and **D**). Expression in WT embryos was stronger than seen in mutant embryos; additionally in 24 hpf WT embryos this stripe of expression was fairly continuous along the length of the zebrafish trunk but in mutants was punctuated with more gaps. In 30 hpf embryos the pattern of *pax7* expression was consistent with that observed in 24 hpf embryos but was restricted to a more posterior trunk region (data not shown). Expression of *pax7* was also seen in cranial NCCs of WT 24 hpf embryos (data not shown). In 30 hpf embryos, *pax7* expression was identified in NCCs migrating solely on the lateral pathway (**Figure 27B** and **D**) but not in mutant embryos (**Figure 27B'**). *Pax7* expressing cells never appeared to be melanising (data not shown). Thus it appears that *pax7* mRNA marks early NCCs of the xanthophore lineage. *Pax7* expression was also detected in dorsal NT (**Figure 27A**, **A'** and **D**) and was weakly expressed in the somites (data not shown). It has been demonstrated that Pax7 positive cells are present on the lateral somite surface and these cells are distinct from NC

derived Pax7 positive cells (Hammond et al., 2007). These non NC sites of expression did not show obvious down-regulation of *pax7* mRNA between WT and *sox10*$^{-/-}$ embryos. *Pax7* has been proposed to be a marker of the long-sought common chromatoblast precursor cell (Lacosta et al., 2007); however no conclusive evidence of *pax7* expression overlapping with chromatoblast markers was provided. A careful study of *pax7* expression in conjunction with markers for the melanocyte and xanthophore lineage concluded that this gene is a marker of xanthoblasts (Minchin and Hughes, 2008). In summary, both the data presented here and published data has established that *pax7* is expressed in xanthoblasts and is down-regulated in *sox10* mutant embryos.

A study of *pax7* expression in chick embryos identified this gene as a very early marker of the NC (Basch et al., 2006). Morpholino mediated inhibition of *pax7* translation resulted in down-regulation of the early NC marker genes *slug*, *sox9* and *sox10* suggesting a role in early NC formation. In addition the number of migrating cranial NCCs was also strongly reduced (Basch et al., 2006). The mouse null *pax7* mutant also shows defects in cranial NC derivatives (Mansouri et al., 1996). Studies in zebrafish have shown that at least some NCCs express *pax7* mRNA and protein (Hammond et al., 2007) and this was later refined to identify *pax7* as a marker of the xanthophore lineage (Minchin and Hughes, 2008). Having conclusively shown that *pax7* marks xanthoblasts, a temporal analysis of *pax7* expression and *xdh* identified that *pax7* expression initiates after *xdh* expression (Minchin and Hughes, 2008). This temporal analysis was suggestive of a role for *pax7* in xanthophore differentiation and a morpholino mediated knockdown approach was applied. *Pax7* morphants displayed a dose dependent reduction in yellow pigmentation but not a reduction in xanthophore cell number confirming a role in differentiation (Minchin and Hughes, 2008). In summary, *pax7* has both clear expression in and a functional role during the development of a NC derivative and is thus a well validated target identified in the *sox10* mutant microarray screen.

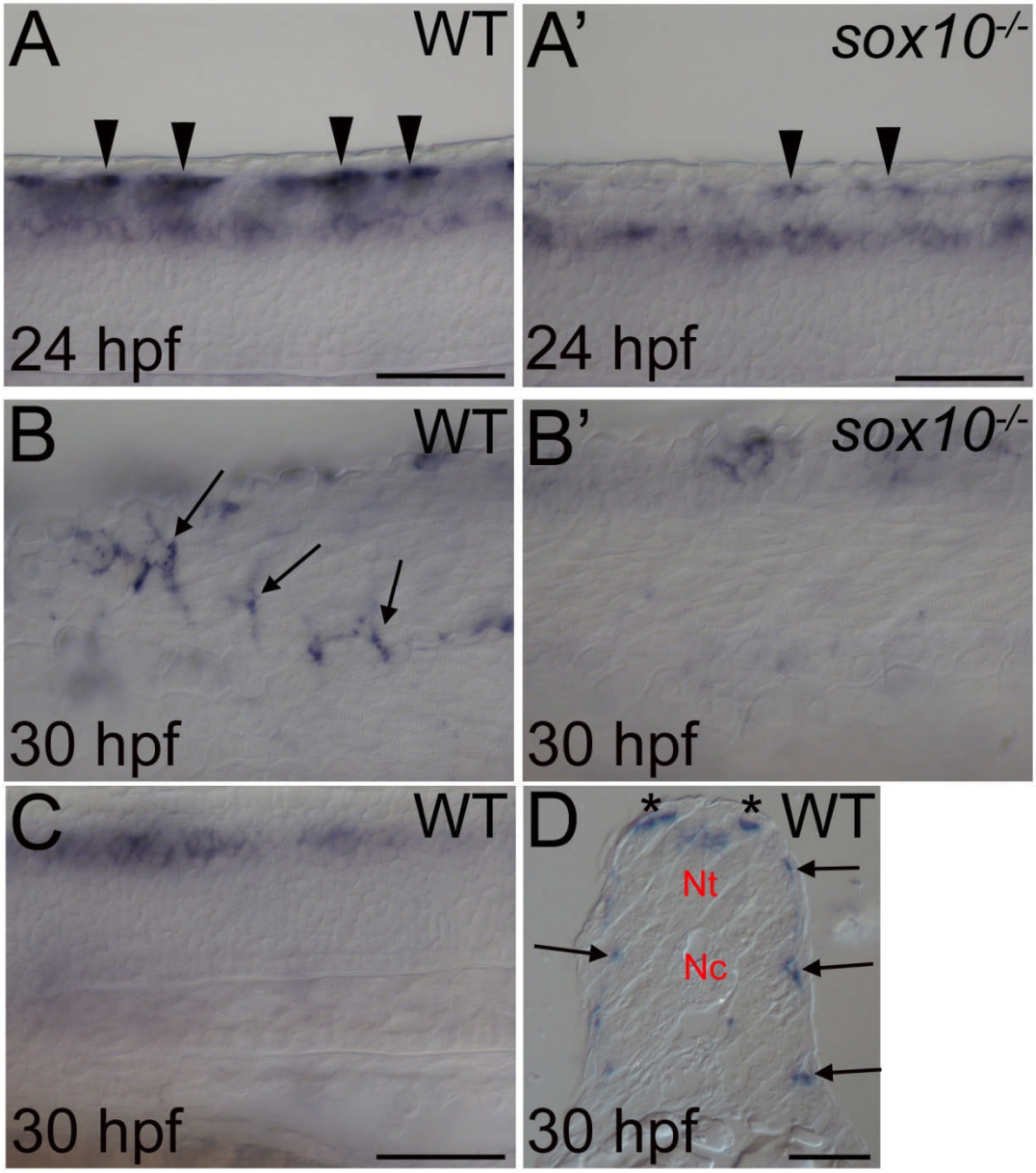


Figure 27: Whole mount *in situ* hybridization pattern of p*ax7 expression*.
Images are of zebrafish trunk and presented as per Figure 20 except image D which is a transverse section through the trunk of a WT 30 hpf embryo with dorsal oriented to the top. Pre-migratory NCCs were strongly labelled in WT 24 hpf (arrowheads in A) and 30 hpf (asterisks in D) embryos, expression in these cells was reduced in mutant embryos (arrowheads in A'). A stripe of expression below the pre-migratory NCCs in the dorsal neural tube appears to be unchanged between WT and *sox10* mutant embryos. In WT 30 hpf embryos cells migrating on the lateral pathway (arrows in B and D) expressed *pax7*, these cells were not detectable in *sox10*$^{-/-}$ embryos (B'). No cells migrating on the medial pathway expressed *pax7* (C). Nt = Neural Tube, Nc = Notochord. Scale bar in D = 25 µm, all other scale bars = 50 µm.

4.2.5.3.3 *slc2a15b*

Affymetrix Probe Identifier		Dr.14747.1.A1_at	Dr.20896.1.S1_at
Affymetrix Gene Symbol		*slc2a15b*	*slc2a15b*
ZFIN Gene Symbol		*slc2a15b*	*slc2a15b*
Full Name		*solute carrier family 2 (facilitated glucose transporter), member 15b*	*solute carrier family 2 (facilitated glucose transporter), member 15b*
Orthology		N/A	N/A
Biological Process GO Term		transport transmembrane transport	transport transmembrane transport
Molecular Function GO Term		transporter activity substrate-specific transmembrane transporter activity	transporter activity substrate-specific transmembrane transporter activity
7.2sox10:GFP array results	*t*-statistic	-9.9877481461	-7.371524334
	p-value	0.0005646773	0.0018048198
	CLEAR result	diff.expr	diff.expr
	Fold Change	-3.03	-1.99
4.9sox10:GFP array results	*t*-statistic	-4.6440958977	-6.865673542
	p-value	0.0433716364	0.0205624737
	CLEAR result	diff.expr	diff.expr
	Fold Change	-3.77	-1.56
7.2+4.9sox10:GFP array results	*t*-statistic	-8.9376220703	-7.8423748016
	p-value	0.0000195028	0.0000503794
	CLEAR result	diff.expr	diff.expr
	Fold Change	-3.46	-1.81

Table 29: Summary of *slc2a15b* information

The Affymetrix probe sets Dr.14747.1.A1_at and Dr.20896.1.S1_at detect transcript levels for the same gene, *slc2a15b*. This gene was identified as down-regulated in all data sets by both statistical tests for both probe sets (**Table 29**). This provided two independent confirmations of differential regulation. Expression of *slc2a15b* was readily detected by *in situ* hybridization in both cranial and trunk NCCs (**Figure 28**). At 24 hpf cranial NCCs migrating just under the epidermis (**Figure 28A** and **B**) and a sub-set of pre-migratory NCCs cells (**Figure 28C**) expressed *slc2a15b* strongly. This expression was strongly reduced, but not absent, in *sox10*$^{-/-}$ embryos which displayed some stained cells in pre-migratory positions (**Figure 28A'** and **C'**). At 30 hpf, some pre-migratory NCCs in the posterior trunk still expressed *slc2a15b* (data not shown). At this stage *slc2a15b* positive cells were migrating extensively over the anterior trunk on the lateral pathway only (**Figure 28D**, **E** and **F**) showing no co-localisation with melanising cells. An expression pattern for this gene has been directly submitted to the ZFIN website (Thisse et al., 2004) but in contrast to **Figure 28**, shows no spatial restriction. The expression patterns revealed here provide strong evidence that *slc2a15b* marks xanthoblasts and is down-

regulated in *sox10* mutants. This gene is extensively expressed in cells of the xanthophore lineage and is likely to prove a useful marker gene.

The gene *slc2a15b* is in the Entrez Gene Database under the ID number 552924. This database shows that *slc2a15b* is part of the major facilitator superfamily, a large and diverse family of transporters. Very little published data exists relating to *slc2a15b*. The *slc2a15b* probe set Dr.14747.1.A1_at was identified by microarray analysis as being over expressed, along with the NC genes *dct, gch* and *crestin*, in the retinal pigmented epithelium when compared to the retina at 52 hpf (Leung et al., 2008). Thus this gene has been the subject of little study and has yet to be assigned a function in NC biology. Based on the nomenclature of this gene it functions as a glucose or sugar transporter, it would be interesting to determine the importance of such a function in NC development. *Slc2a15b* provides an example of a novel gene identified by this screen.

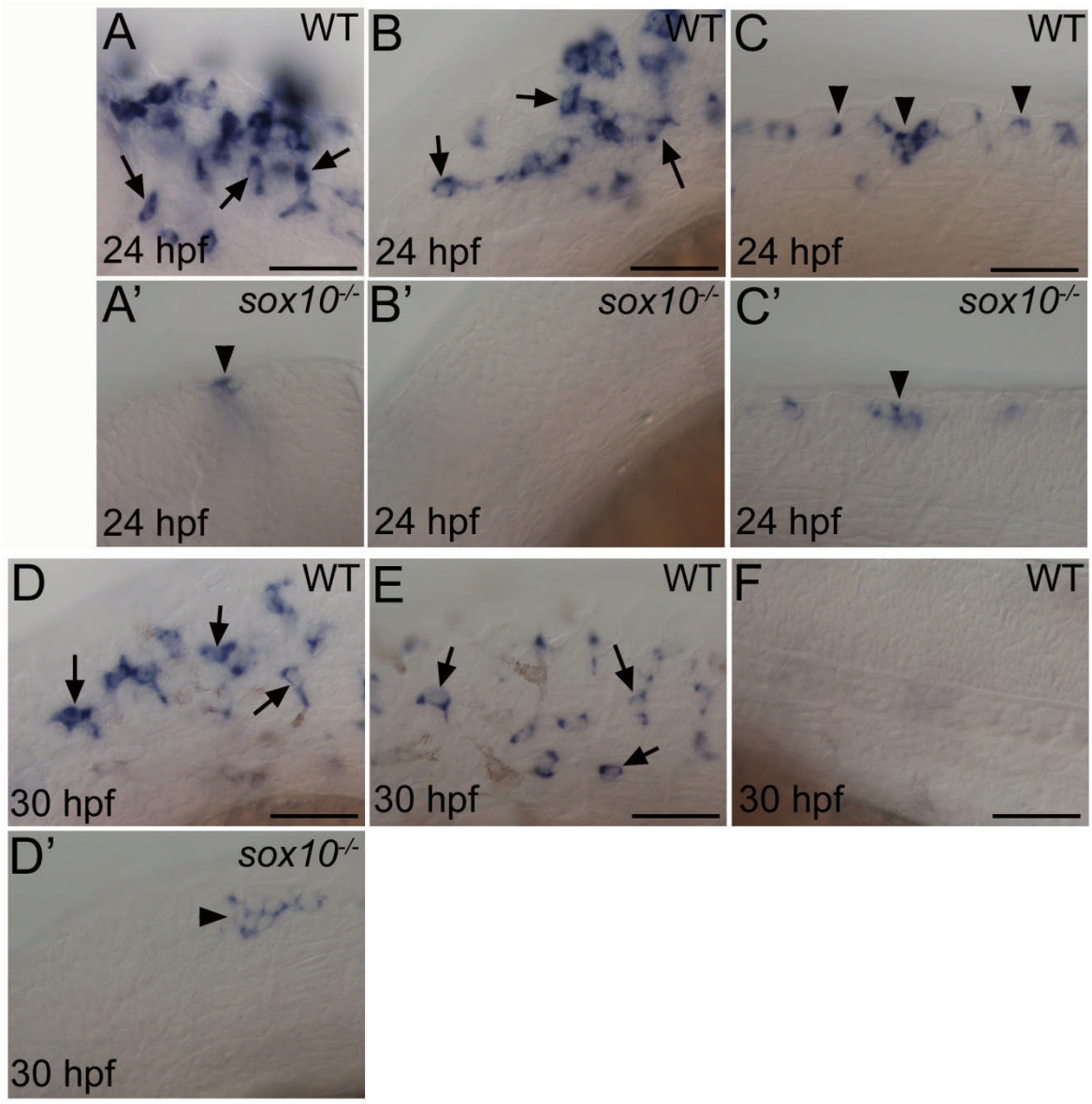


Figure 28: Whole mount *in situ* hybridization pattern of *slc2a15b* expression.
Images are presented as per Figure 20. Migrating cells are indicated by arrows, cells in pre-migratory positions are indicated by arrowheads. Images A and A' are of the head at the midbrain-hindbrain boundary, images B, B', D and D' are of posterior head and anterior trunk. Remaining images are of mid-trunk. Migrating NCCs are indicated with arrows, pre-migratory NCCs are indicated by arrowheads. Scale bars = 50 µm.

4.2.5.3.4 *cx33.8*

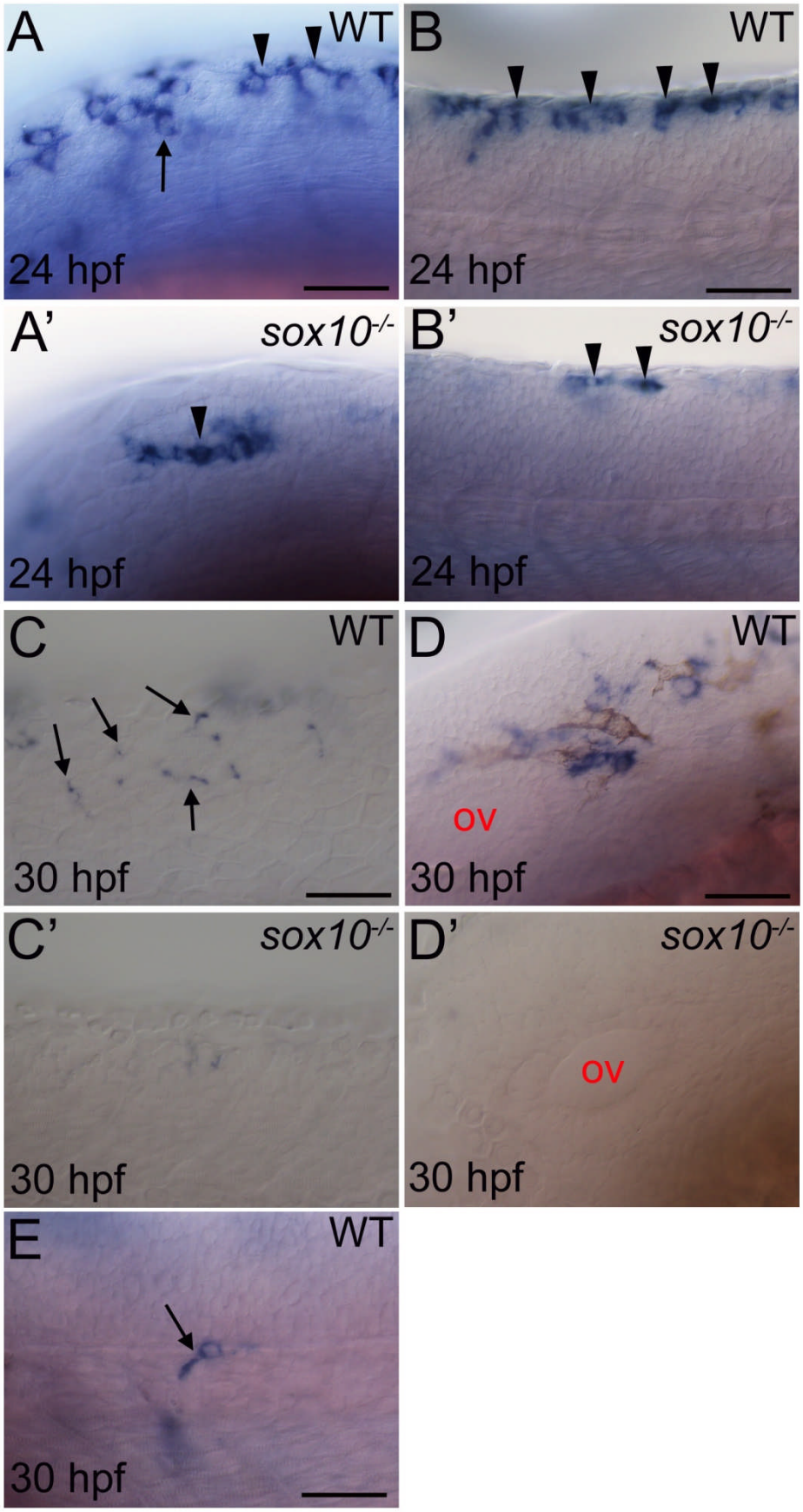
Affymetrix Probe Identifier		DrAffx.1.65.S1_at
Affymetrix Gene Symbol		*cx33.8*
ZFIN Gene Symbol		*cx33.8*
Full Name		*connexin 33.8*
Orthology		*GJB2* (*CX26*) and *GJB6* (*CX30*)
Biological Process GO Term		cell communication
Molecular Function GO Term		N/A
7.2sox10:GFP array results	*t*-statistic	-8.6883687973
	p-value	0.0009660177
	CLEAR result	diff.expr
	Fold Change	-1.54
4.9sox10:GFP array results	*t*-statistic	-3.7531321049
	p-value	0.0642285049
	CLEAR result	diff.expr
	Fold Change	-1.64
7.2+4.9sox10:GFP array results	*t*-statistic	-5.5526852608
	p-value	0.0005392728
	CLEAR result	diff.expr
	Fold Change	-1.56

Table 30: Summary of *cx33.8* information.

The Affymetrix probe set DrAffx.1.65.S1_at corresponds to the gene *cx33.8*. This gene was detected as differentially regulated by both statistical tests in all the data sets except by the *t*-Test on the 4.9 data set (**Table 30**). A strong NC gene expression pattern was identified by *in situ* hybridization for *cx33.8* (**Figure 29**). At 24 hpf strong pre-migratory NC expression was observed (arrowheads in **Figure 29A** and **B**) and some cells migrating on the lateral pathway at the anterior-trunk axial level were also noted (arrow **Figure 29A**). These sites of expression were strongly reduced in mutant embryos although some clustered pre-migratory cells continued to express *cx33.8* (arrow heads in **Figure 29A'** and **B'**). Composites showing the expression pattern of *cx33.8* in both WT and *sox10* mutant 24 hpf embryos are also shown for comparison displaying the features noted above (**Figure 29F** and **F'**). At 30 hpf, in WT embryos, *cx33.8* expression remained in pre-migratory NCCs in the posterior trunk (data not shown). Cells weakly expressing *cx33.8* were visible migrating on the lateral pathway in WT embryos but not in mutant embryos (**Figure 29C** and **C'**). Rarely *cx33.8* expressing cells were seen migrating on the medial pathway; in approximately 50 % of the WT embryos a single cell was typically observed. (**Figure 29E**). *Cx33.8* expressing cells did not appear to melanise (**Figure 29D**). *Cx33.8* expression was strongest at 24 hpf and therefore is likely to be expressed in early cells of the xanthophore lineage. The pattern of *cx33.8* expression at 24 hpf closely matches that of *xdh* and *gch* (see Chapter 3 and Pelletier et al., 2001). *Cx33.8* has also been shown to be expressed in the otic vesicle of zebrafish embryos at 3 dpf and is down-regulated at this site in *sox10*$^{-/-}$ embryos (Dutton et al., 2009). In summary *cx33.8* appears to be expressed chiefly in xanthoblasts during early embryonic development with expression in

unidentified occasional medial pathway migrating cells also observed. Thus for the first time, *cx33.8* has been identified as a marker of a sub-set of NCCs including xanthoblasts and this expression is reduced in *sox10* mutant embryos.

The zebrafish gene *cx33.8* is an orthologue of two human genes, GJB2 (CX26) and GJB6 (CX30) (Cruciani and Mikalsen, 2006). Mutations in these two genes that result in deafness have been characterised in a number of populations (Rabionet et al., 2002). Thus *cx33.8* expression in the zebrafish ear suggests at a conserved role in ear development. However the data presented here provides the first evidence for a role in neural crest biology. Little is known about the *cx33.8* gene in zebrafish thus making this target an attractive one to investigate further to elucidate a function in neural crest and xanthophore biology (see discussion).


A
WT
24 hpf
B
WT
24 hpf
A'
sox10-/-
24 hpf
B'
sox10-/-
24 hpf
C
WT
30 hpf
D
WT
ov
30 hpf
C'
sox10-/-
30 hpf
D'
sox10-/-
ov
30 hpf
E
WT
30 hpf

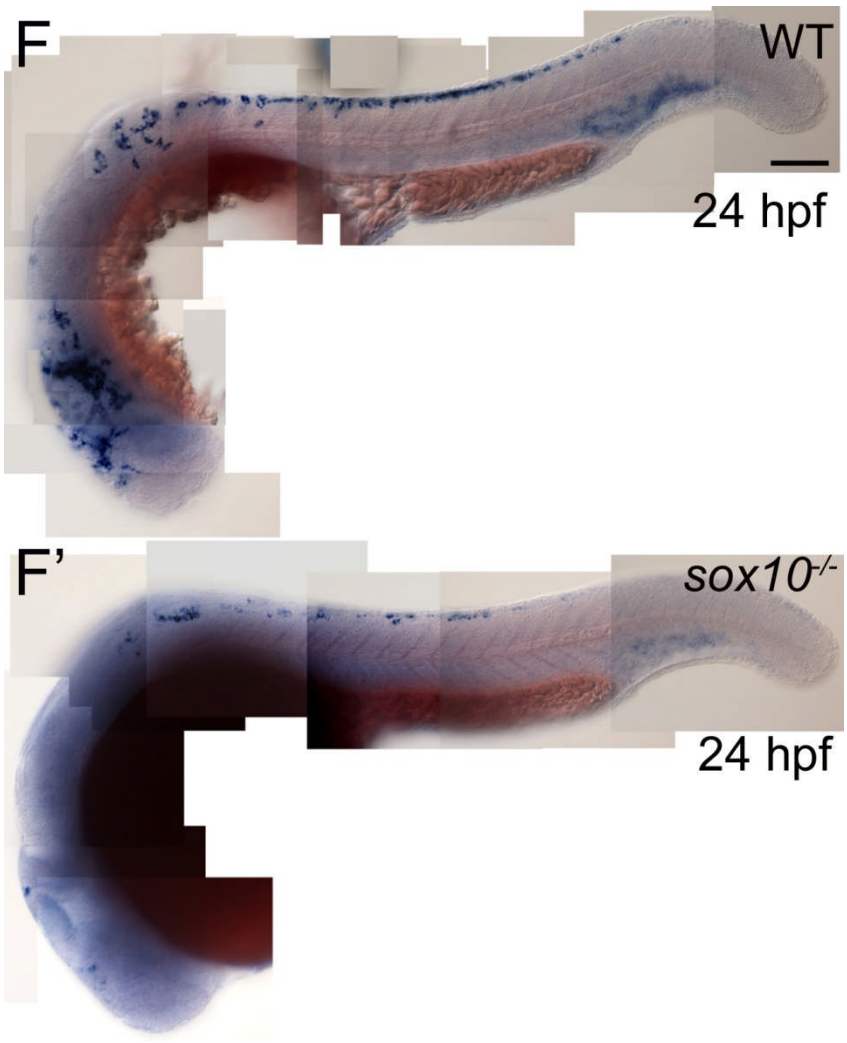


Figure 29: Whole mount *in situ* hybridization pattern of *cx33.8* expression.
All images are presented as per Figure 20 except images F and F' which are composite images in the same orientation. Images A and A' are of anterior trunk, images B-C' and E are of mid-trunk and images D and D' are of posterior head including the otic vesicle (ov). Pre-migratory NCCs are marked by arrowheads (A-B') and migratory NCCs are highlighted with arrows (A, C and E). Scale bars in F and F' = 100 µm, all other scale bars = 50 µm.

4.2.5.3.5 *Validation of three genes in the IMP biosynthesis pathway*

Affymetrix Probe Identifier		Dr.481.1.S1_at	Dr.983.1.S1_at	Dr.13359.1.S1_at
Affymetrix Gene Symbol		*atic*	*paics*	*adsl*
ZFIN Gene Symbol		*atic*	*paics*	*adsl*
Full Name		5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase	phosphoribosylaminoimidazole carboxylase, phosphoribosylaminoimidazole succinocarboxamide synthetase	adenylosuccinate lyase
Orthology		*ATIC*	*PAICS*	*ADSL*
Biological Process GO Term		purine nucleotide biosynthetic process IMP biosynthetic process	purine nucleotide biosynthesis 'de novo' IMP biosynthesis	purine ribonucleotide biosynthetic process
Molecular Function GO Term		catalytic activity IMP cyclohydrolase activity phosphoribosylaminoimidazolecarboxamide formyltransferase activity transferase activity hydrolase activity	DNA binding catalytic activity phosphoribosylaminoimidazole carboxylase activity phosphoribosylaminoimidazolesuccinocarboxamide synthase activity ATP binding	catalytic activity adenylosuccinate lyase activity lyase activity
7.2sox10:GFP array results	*t*-statistic	-8.6674394608	-7.7576050758	-7.0870451927
	p-value	0.0009749859	0.0014879868	0.0020928441
	CLEAR result	diff.expr	diff.expr	diff.expr
	Fold Change	-1.85	-1.56	-1.84
4.9sox10:GFP array results	*t*-statistic	-3.0787854195	-6.7212591171	-2.9042654037
	p-value	0.0912824571	0.0214270819	0.10092666
	CLEAR result	diff.expr	diff.expr	diff.expr
	Fold Change	-1.80	-2.11	-1.71
7.2+4.9sox10:GFP array results	*t*-statistic	-7.9875617027	-8.1890239716	-5.87099123
	p-value	0.00004416	0.000036897	0.0003737883
	CLEAR result	diff.expr	diff.expr	diff.expr
	Fold Change	-1.88	-1.76	-1.82

Table 31: Summary of information for *atic, paics* and *adsl*.

The Affymetrix probe sets Dr.481.1.S1_at, Dr.983.1.S1_at and Dr.13359.1.S1_at correspond to the genes *atic*, *paics* and *adsl* respectively (**Table 31**). All three genes encode for enzymes that function in the IMP biosynthesis pathway (**Figure 33**). These genes were all identified as down-regulated in the *7.2sox10*:GFP microarray data set by the *t*-Test and the CLEAR test. Whole mount *in situ* hybridization was used to validate the microarray results for *paics* (**Figure 30**) *atic* (**Figure 31**) and *adsl* (**Figure 32**).

Whole mount *in situ* hybridization revealed that *paics* had a strong NC expression pattern (**Figure 30**). At 24 hpf *paics* was expressed in a sub-set of pre-migratory NCCs along much of the trunk and in cells migrating on the lateral pathway in anterior trunk (**Figure 30A, B** and **C**). Expression at pre-migratory NC sites was strongly reduced in *sox10*$^{-/-}$ embryos and no cells were observed migrating on the lateral pathway (**Figure 30A'** and **B'**). *Paics* expression was also observed in cranial NCCs and this was reduced in mutant embryos (data not shown). A similar expression pattern was noted at 30 hpf although pre-migratory NCCs were restricted to a more posterior region and lateral pathway migrating cells had extended into the mid-trunk region (**Figure 30D, E** and **F**). Again pre-migratory NCC expression was reduced and migratory cells were absent in mutant embryos (**Figure 30D', E'** and **F'**). Expression of *paics* did not overlap with melanising cells (**Figure 30D**). The NC expression pattern of *paics,* including cells migrating on the lateral pathway, has already been published on the ZFIN website (Thisse et al., 2004). Thus this gene appears to be expressed in early cells of the xanthophore lineage, is down-regulated in *sox10* mutant embryos and with a strong expression pattern may prove to be a useful marker. Consistent with this, an insertion mutant (*paics*^{hi2688}) has been characterised for this gene. This mutant has a lack of pigmentation, although it is unclear if this is due to a loss of xanthophores or a failure in pigmentation. Melanocytes appear to be grossly normal (Amsterdam et al., 2004). Thus the evidence presented here (**Figure 30**) and other published evidence suggests that *paics* is expressed in xanthoblasts, is down-regulated in *sox10* mutant embryos and plays an important role in xanthophore development.

The expression patterns of *atic* (**Figure 31**) and *adsl* (**Figure 32***)* were both weaker than, but displayed many similarities to, the expression of *paics*. Essentially, expression of both genes was noted in typical NC positions and in cells migrating on the lateral pathway; these patterns of expression were absent in *sox10* mutant embryos. However, some differences were noticed and these will be explored here. The expression pattern of *atic* was very weak and therefore difficult to interpret. Only a few cells were noted migrating on the lateral pathway (**Figure 31F**), but no cells were observed migrating on the medial pathway (data not shown). *Atic* expression also appeared to mark fewer cells than *paics. Adsl* expression was also less readily detected than *paics* expression and thus fewer cells appeared to be marked. This may be a result of the decreased sensitivity of detecting *atic* or *adsl* expression or a result of a more restricted temporal period of expression for these genes. Rare *adsl* positive cells were observed migrating on the medial pathway in WT 24 hpf embryos but not *sox10* mutant embryos (**Figure 32C** and **C'**). *Adsl* expression in medial pathway migrating cells was only observed in one or two cells per embryos. While these cells appear to be NCCs, it is unknown what cell type they are. At 30 hpf in both WT and *sox10* mutant embryos displayed strong *adsl* expression in

the somites (**Figure 32E** and **E'**). This made it very difficult to identify if expression in medial pathway migrating NCCs was maintained at this stage. The *atic* expression pattern has also been published on the ZFIN website showing expression in NCCs migrating on the lateral pathway at 24 hpf (Thisse et al., 2004). No embryonic expression pattern has been published for *adsl* in zebrafish. Despite the reduced number of cells displaying expression of these two genes, the expression patterns shown here are consistent with expression in cells of the xanthophore lineage. These gene expression patterns also show strong similarities to the expression pattern of *paics*, a gene in the same metabolic pathway (**Figure 33**).

While examining the published expression patterns of both *paics* and *atic*, it was observed that expression was visible in a subset of NCCs at 2 dpf (Thisse et al., 2004). These cells were located in the dorsal stripe and in part of the ventral stripe. The location of these cells and their interspersed distribution are consistent with the iridophores lineage. As no embryonic pattern of *adsl* expression has been published it would be worth examining 2 dpf embryos for a similar expression pattern.

The enzyme encoded by the *paics* gene is a bifunctional enzyme that catalyses steps 6 and 7 in the *de novo* purine biosynthesis pathway (**Figure 33**) (Li et al., 2007). This enzyme has both 5-aminoimidazole ribonucleotide (AIR) carboxylase (AIRC) and 4-(N-succinylcarboxamide)-5-aminoimidazole ribonucleotide (SAICAR) synthetase (SAICARs) activities (Li et al., 2007). *Adsl* encodes the third enzyme in the *de novo* IMP biosynthesis pathway that was validated during the *in situ* hybridization screen. This enzyme catalyses step 8 and performs the conversion of succinylaminoimidazole carboxamide ribotide (SAICAR) into aminoimidazole carboxamide ribotide (AICAR). This is the step between those catalysed by *paics* and *atic* (**Figure 33**). The *ATIC* gene is also known as *PURH, AICAR, AICARFT,* and *IMPCHASE*. This gene encodes a bifunctional enzyme that has both Aminoimidazole ribonucleotide formyltransferase (AICARFT) and IMP cyclohydrolase activity. As a result of this, ATIC catalyses the final two steps in the *de novo* purine biosynthetic pathway (Rayl et al., 1996). The results presented in this section suggest that IMP biosynthesis is involved in xanthophore development. This pathway produces inosine monophosphate (IMP) (**Figure 33**) and other purine nucleotides are then synthesised from this intermediate. The possible significance of this pathway is examined in more detail in the discussion.

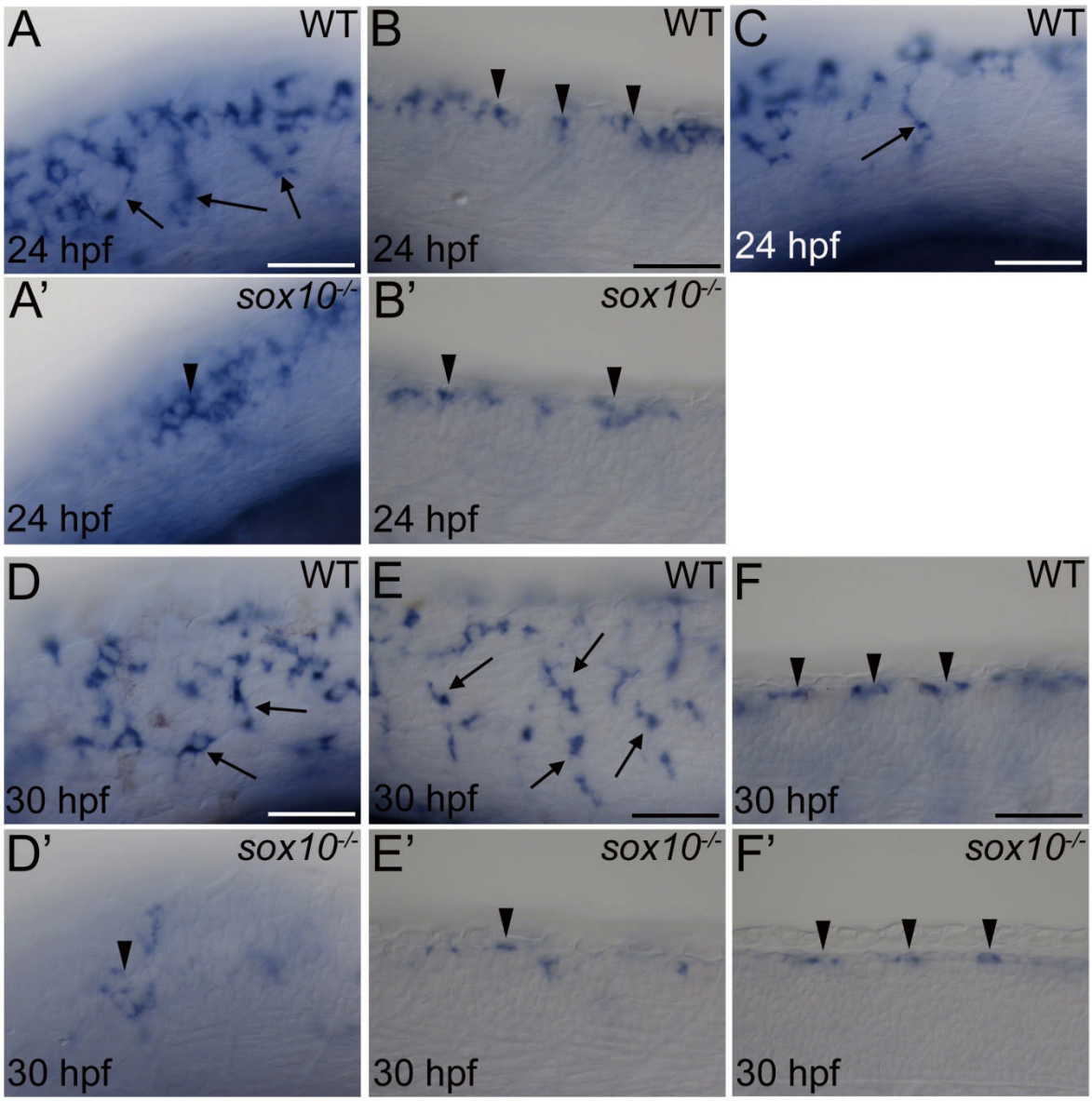


Figure 30: Whole mount *in situ* hybridization pattern of *paics* expression.
All images are presented as per Figure 20. Images are of the following axial levels; A, A', C, D and D' are of anterior trunk, images B, B', E and E' are of mid-trunk and F and F' are of posterior trunk. Arrows indicate migratory cells on the lateral pathway, arrowheads indicate pre-migratory cells. All scale bars = 50 µm.

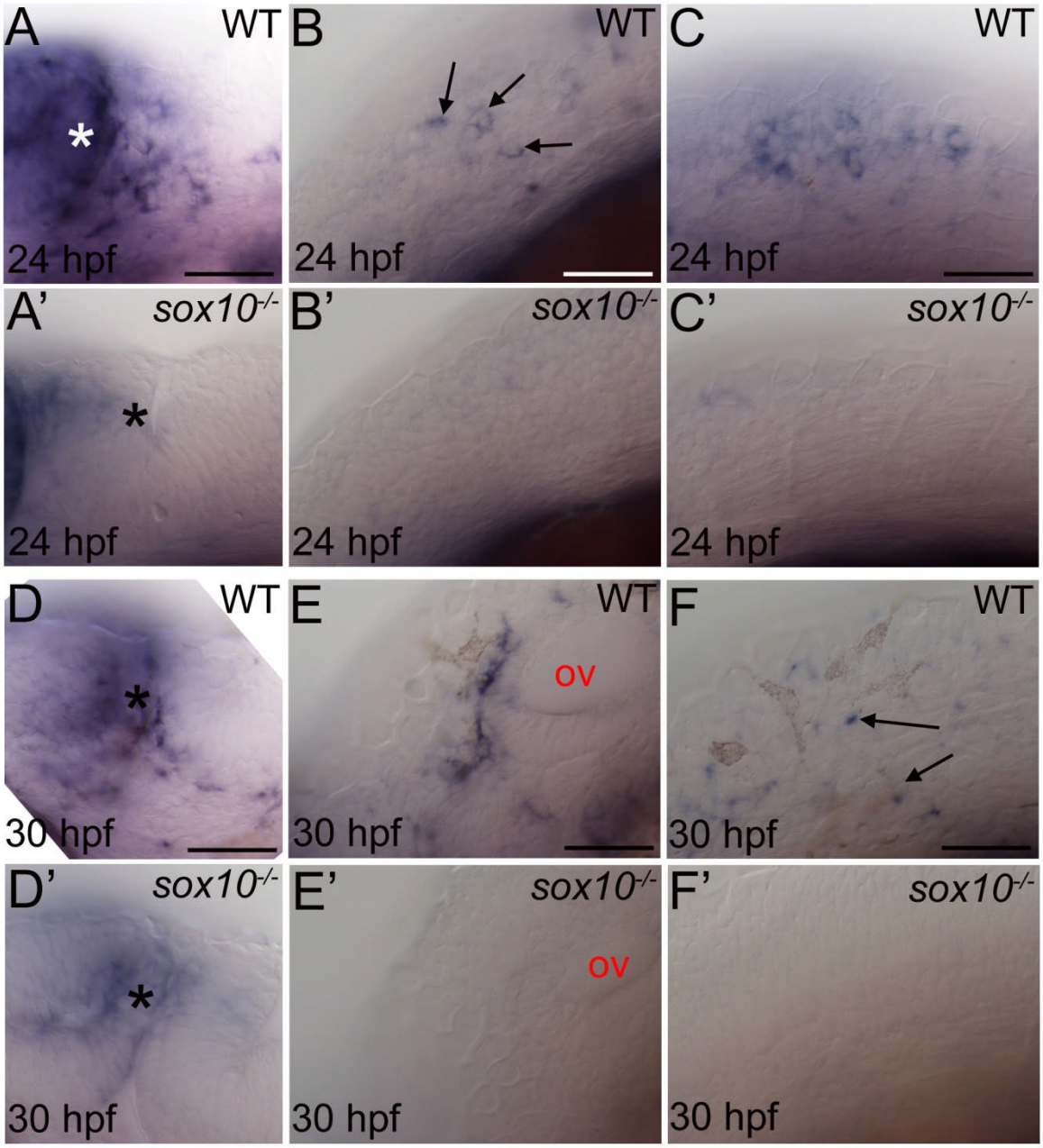


Figure 31: Whole mount *in situ* hybridization pattern of *atic* expression.
All images are presented as per Figure 20. NCC cells expressing *atic* were observed migrating in the midbrain-hindbrain boundary region (* A and D), but not in mutant embryos (A' and D'). *Atic* positive cells migrating on the lateral pathway in anterior trunk (arrows in B and F) were detected in WT embryos but were absent in mutant embryos (B' and F'). Expression in NCCs was also detected in the anterior trunk (C) of WT embryos but not in mutants (C'). *Atic* expressing cells did not overlap with melanising cells (E and F). Ov = otic vesicle, all scale bars = 50 µm.

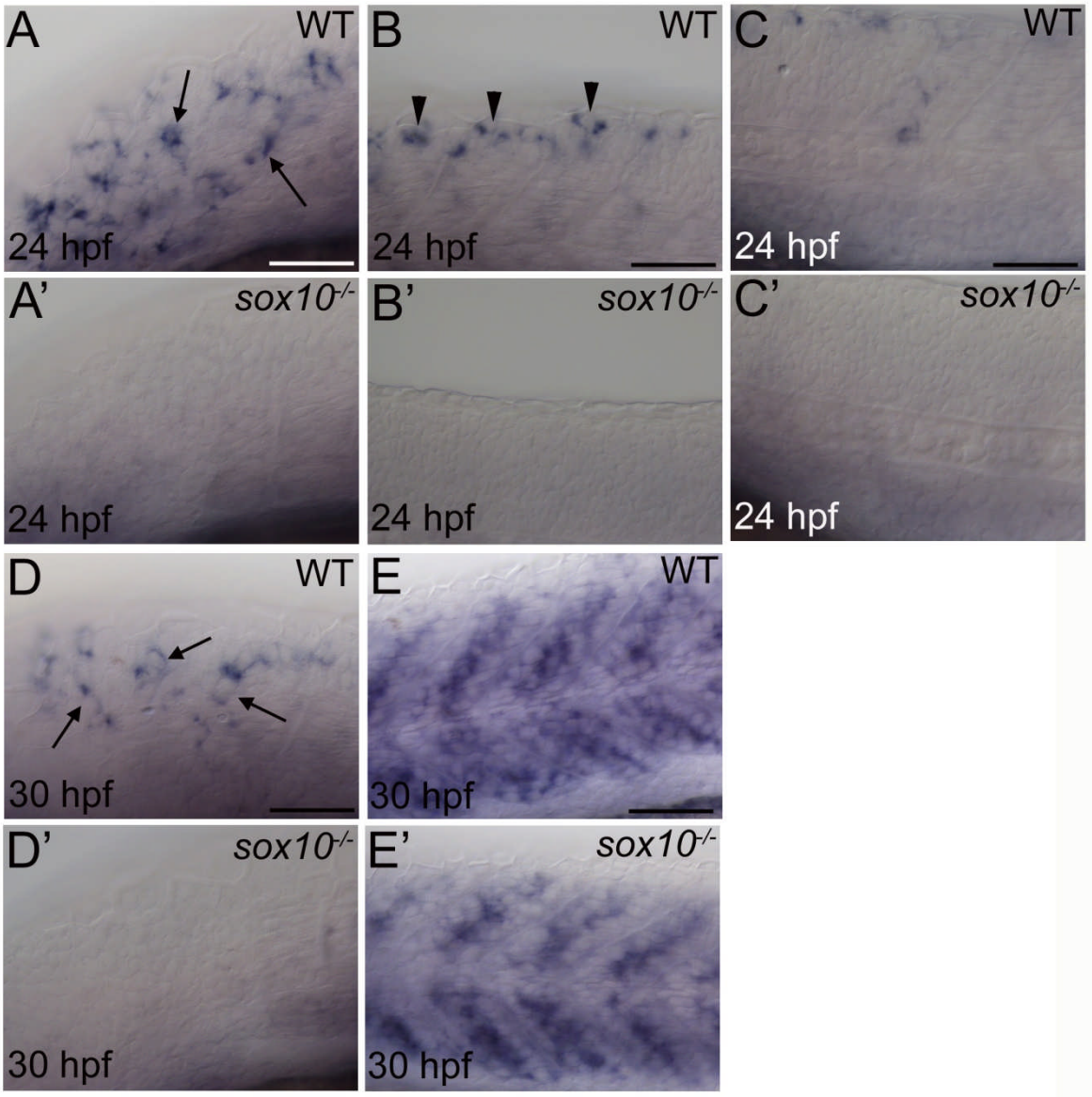


Figure 32: Whole mount *in situ* hybridization pattern of *adsl* expression.
All images are lateral views presented as per Figure 20. Anterior trunk is shown in images A, A', D and D'; images B-C', E and E' show mid-trunk. *Adsl* expression was detected in lateral pathway migrating cells at 24 hpf (arrows in A) and 30 hpf (arrows in D), no expression was detected at these sites in mutant embryos (A' and D'). Rare cells were observed expressing *adsl* migrating on the medial pathway in WT embryos (C) but not mutant embryos (C') at 24 hpf. Expression was also detected in WT pre-migratory NCCs (arrowheads in B) but not in mutant embryos (B'). A strong somite expression pattern was identified at 30hpf (D and D'). All scale bars = 50 µm.

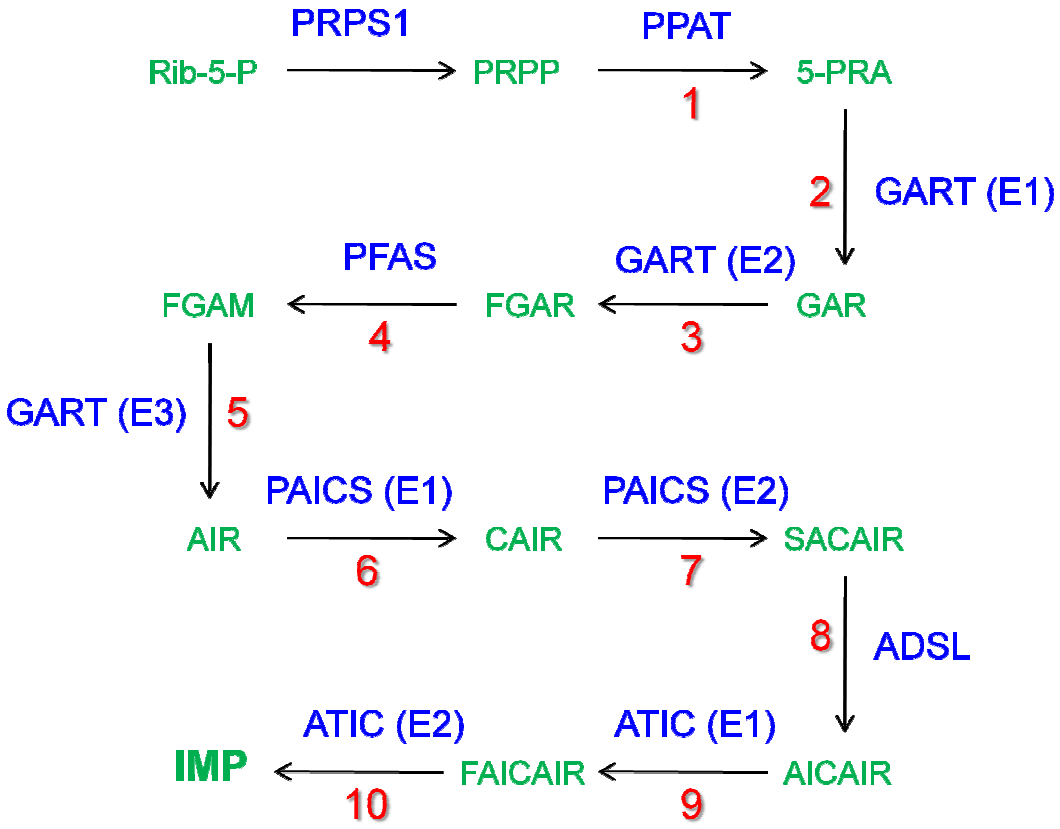


Figure 33: Schematic of the *de novo* IMP biosynthesis pathway.
Enzyme names are in blue, substrate names in green and numbered steps of the pathway are given in red. Where an enzyme acts in more than one step this is indicated. Also see (Christopherson et al., 2002). Rib-5-P = ribose-5-phosphate, PRPP = phosphoribosyl pyrophosphate, 5-PRA = phosphoribosylamine, GAR = phosphoribosyl glycineamide, FGAR = 5-phosphoribosyl-N-formylglycineamidine, FGAM = 5-phosphoribosyl-N-formylglycineamidine, AIR = 5-aminoimidazole ribonucleotide, CAIR = 4-carboxy aminoimidazole ribonucleotide, SACAIR = N-succinyl-5-aminoimidazole-4-carboxamide ribonucleotide, AICAIR = 5-aminoimidazole-4-carboxamide ribonucleotide, FAICAIR = 5-formaminoimidazole-4-carboxamide ribonucleotide and IMP = inosine 5'-monophosphate.

Three genes encoding for enzymes in the *de novo* IMP biosynthesis pathway have been identified as down-regulated in *sox10* mutant embryos by the microarray screen and validated by *in situ* hybridization. The performances of the remaining genes in the pathway were therefore examined in the three microarray data sets (**Table 32**). It was predicted that these genes would be down-regulated in mutant embryos. Two *prps1* genes have been identified in zebrafish; *prps1a* was identified as down-regulated by the *t*-Test in the 4.9 and 7.2+4.9 data sets but the microarray analysis did not indicate that *prps1b* was down-regulated. Neither of these genes appear to be expressed in the NC at 24 hpf (Thisse et al., 2004) but *prps1a* is expressed in pigment cells at 2 dpf in a position consistent with the iridophore lineage (Thisse et al., 2004). It would be worth repeating these tests and comparing results between WT and *sox10* mutant embryos to further assess the expression pattern of these genes. Both *ppat* and *gart* were consistently

identified as down-regulated by both statistical tests in all three data sets. *Ppat* and *gart are* expressed in NCCs migrating on the lateral pathway at 24 hpf (Thisse et al., 2004). Additionally *gart* is expressed in a putative iridophore position at 2 dpf while no *ppat* expression was detected in an iridophore location (Thisse et al., 2004). An insertional mutant for *gart* has been characterised (*gart*^{hi3526b}) that displays a defect in yellow pigmentation and has a strong similarity to the *paics* insertion mutant. No Affymetrix probe set has been attributed to the *pfas* gene and there is no published expression pattern for this gene currently so it could not be examined. In summary, this evidence indicates that the majority of genes in the *de novo* IMP biosynthesis pathway are likely to be down-regulated in *sox10* mutant embryos and are likely to play a role in xanthophore biology during early embryo development.

		prps1a Dr.2896.1.S1_at	*prps1b* Dr.7674.1.S1_at	*ppat* Dr.26358.1.A1_at	*gart* Dr.4801.1.S1_at
7.2sox10 :GFP array results	*t*-statistic	-1.56	-0.23	-5.03	-5.27
	p-value	0.19	0.83	0.0073	0.0062
	CLEAR result	non.sig	non.sig	diff.expr	diff.expr
	Fold Change	-1.04	-1.03	-1.62	-2.07
4.9sox10 :GFP array results	*t*-statistic	-4.33	-2.09	-5.46	-11.39
	p-value	0.049	0.17	0.032	0.0076
	CLEAR result	non.sig	non.sig	diff.expr	diff.expr
	Fold Change	-1.06	-1.39	-1.54	-1.77
7.2+4.9 sox10 :GFP array results	*t*-statistic	-3.016	-1.20	-5.39	-5.16
	p-value	0.017	0.26	0.00066	0.00086
	CLEAR result	non.sig	non.sig	diff.expr	diff.expr
	Fold Change	-1.05	-1.14	-1.60	-1.91

Table 32: Microarray Results for the remaining *de novo* purine biosynthesis pathway genes.

4.2.5.3.8 Dr.10624 (*zgc:110343*)

Affymetrix Probe Identifier		Dr.10624.2.S1_a_at	Dr.10624.2.S1_at
Affymetrix Gene Symbol		*zgc:110343*	*zgc:110343*
ZFIN Gene Symbol		*zgc:110343*	*zgc:110343*
Full Name		*zgc:110343*	*zgc:110343*
Orthology		*PRDX2*	*PRDX2*
Biological Process GO Term		cell redox homeostasis	cell redox homeostasis
Molecular Function GO Term		antioxidant activity oxidoreductase activity	antioxidant activity oxidoreductase activity
7.2sox10:GFP array results	*t*-statistic	-7.7046275139	-8.217423439
	p-value	0.0015271413	0.0011953737
	CLEAR result	diff.expr	diff.expr
	Fold Change	-2.84	-2.08
4.9sox10:GFP array results	*t*-statistic	-20.3884429932	-15.4375991821
	p-value	0.0023970008	0.0041698213
	CLEAR result	diff.expr	diff.expr
	Fold Change	-3.59	-3.10
7.2+4.9sox10: GFP array results	*t*-statistic	-8.1723003387	-6.6160578728
	p-value	0.0000374464	0.0001665774
	CLEAR result	diff.expr	diff.expr
	Fold Change	-3.18	-2.48

Table 33: Summary of information for *zgc:110343.*

Both the Affymetrix probe sets Dr.10624.2.S1_a_at and Dr.10624.2.S1_at correspond to the gene *zgc:110343*. Both probe sets were identified as down-regulated in all three data sets by both statistical sets, furthermore a third probe set that also corresponds to this gene showed similar results but was outside the bottom 100 most down-regulated genes. Whole mount *in situ* hybridization was performed to identify the expression pattern of this gene (**Figure 34**). Strong *zgc:110343* expression was seen at 24 hpf in WT cranial NCCs, in a subset of pre-migratory trunk NCCs and in NCCs migrating on the lateral pathway in anterior trunk (**Figure 34A**, **B** and **C**). Expression at these sites was strongly reduced in *sox10* mutant embryos and cells failed to migrate in mutants resulting in clumps of cells remaining in pre-migratory positions (**Figure 34A'**, **B'** and **C'**). In 30 hpf WT embryos a similar expression pattern was observed with more extensive migration on the lateral pathway (**Figure 34D**). No migration on the medial pathway was observed (**Figure 34E**). In 30 hpf *sox10* mutant embryos a few cells still expressed *zgc:110343* in pre-migratory positions similar to 24 hpf (**Figure 34D'**). The expression pattern of *zgc:110343* shows the features of an early xanthophore lineage marker with cells only migrating on the lateral pathway, not overlapping with melanocytes and some expression remaining in pre-migratory clumps of cells in *sox10*$^{-/-}$ embryos. This gene has an expression pattern published on ZFIN which shows a very similar pattern as shown in **Figure 34** although this published pattern does show more extensive *zgc:110343* expression (Thisse et al., 2004). In summary, this gene shows strong expression in xanthoblasts at both 24 hpf and 30 hpf,

is likely to be an excellent marker for this cell type and is down-regulated in *sox10* mutants.

The human orthologue of *zgc:110343* was annotated as *Peroxiredoxin 2* (*PRDX2*), although a *prdx2* gene has already been identified in zebrafish. The zebrafish *prdx2* gene is not expressed in the NC (Thisse et al., 2004) and has also been identified as an orthologue of the human *PRDX2* gene. This suggests that *prdx2* has been duplicated in zebrafish. Peroxiredoxin genes play an antioxidant protective role in both normal cells types and cancerous cells (Shen and Nathan, 2002, Manandhar et al., 2009). A major role of peroxiredoxins is the enzymatic degradation of hydrogen peroxide to prevent it being converted into reactive oxygen species (Immenschuh and Baumgart-Vogt, 2005). Hydrogen peroxide can also act as an intracellular messenger and thus peroxiredoxins can also mediate this signalling pathway (Kang et al., 1998, Kang et al., 2004). High levels of peroxiredoxins mediating hydrogen peroxide signalling have been associated with both cell proliferation and inhibition of apoptosis (Immenschuh and Baumgart-Vogt, 2005). Thus a peroxiredoxin gene such as *zgc:110343* may play a role in preventing oxidative stress in NCCs or in the regulation of proliferation or cell survival. However no definitive role for *prdx2* or *zgc:110343* has been characterised during NC development thus making this an interesting gene for further study.

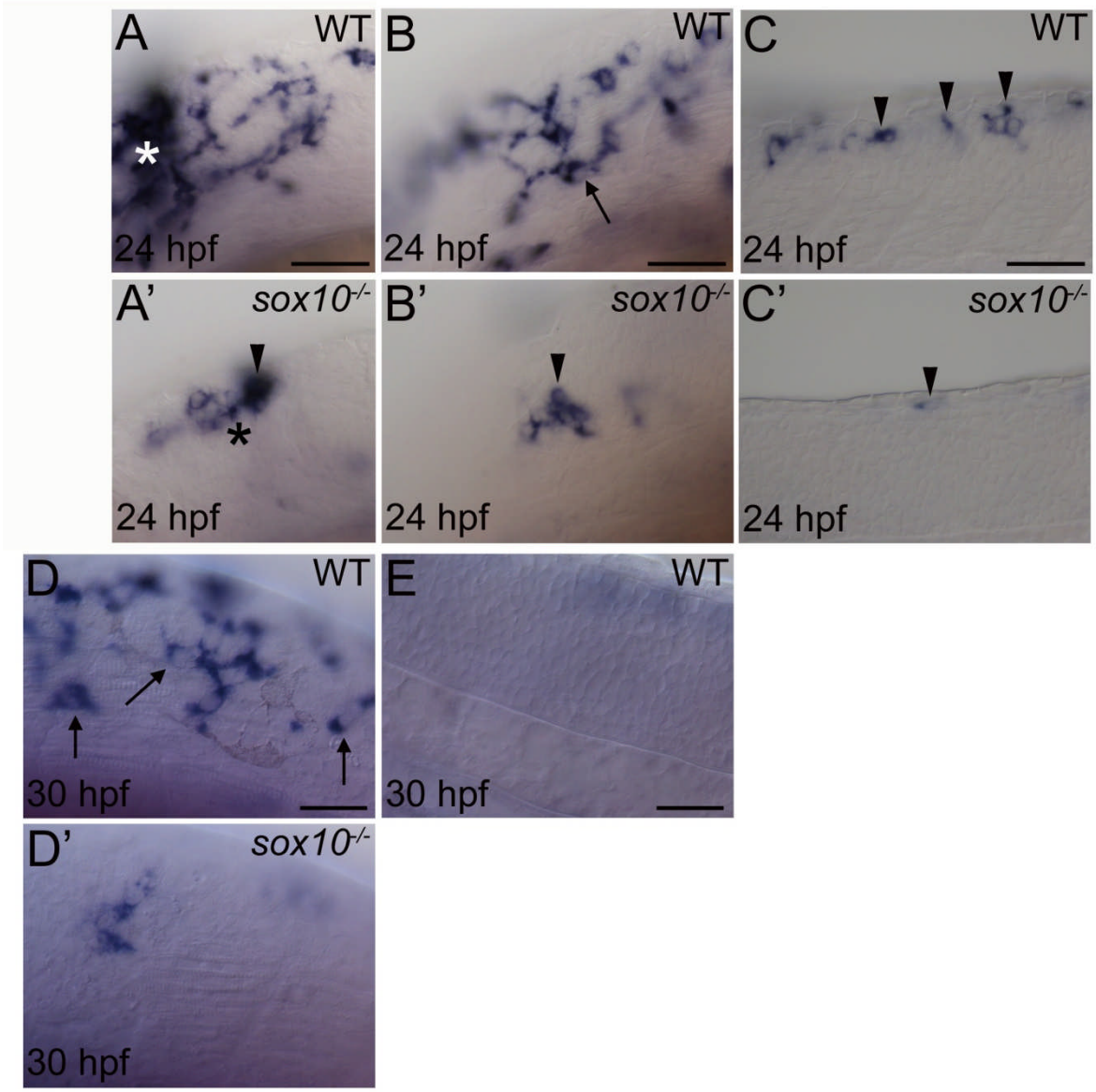


Figure 34: Whole mount *in situ* hybridization pattern of *zgc:110343* expression.
All images are presented as per Figure 20. *Zgc:110343* was expressed in cranial NCCs (A) migrating around the midbrain-hindbrain boundary (*) and in cells migrating on the lateral pathway (arrows) in anterior trunk at both 24 hpf (B) and 30 hpf (D). In mid trunk at 24 hpf cells expressing *zgc:110343* were present in pre-migratory positions (arrowhead in C). At all these sites in *sox10* mutant embryos expression was strongly reduced although some cells expressing *zgc:110343* remained in pre-migratory positions (arrowheads in A', B', C' and D'). *Zgc:110343* expression did not overlap with melanocytes (D) or appear on the medial pathway (E). Scale bars in A-C = 50 µm, scale bars in D and E = 25 µm.

4.2.5.3.9 aldh2

Affymetrix Probe Identifier		Dr.4751.1.S1_a_at	Dr.4751.2.A1_at
Affymetrix Gene Symbol		*aldh2a, aldh2b*	*aldh2b*
ZFIN Gene Symbol		*aldh2a, aldh2b*	*aldh2b*
Full Name		*aldehyde dehydrogenase 2a/b*	*aldehyde dehydrogenase 2b*
Orthology		*ALDH2*	*ALDH2*
Biological Process GO Term		cellular aldehyde metabolic process metabolic process	cellular aldehyde metabolic process metabolic process
Molecular Function GO Term		nucleic acid binding 3-chloroallyl aldehyde dehydrogenase activity zinc ion binding oxidoreductase activity metal ion binding	nucleic acid binding 3-chloroallyl aldehyde dehydrogenase activity zinc ion binding oxidoreductase activity metal ion binding
7.2sox10:GFP array results	*t*-statistic	-7.7480130196	-6.883207798
	p-value	0.0014949832	0.0023346485
	CLEAR result	diff.expr	diff.expr
	Fold Change	-2.34	-2.48
4.9sox10:GFP array results	*t*-statistic	-5.3127570152	-9.3952064514
	p-value	0.0336509645	0.0111399386
	CLEAR result	diff.expr	diff.expr
	Fold Change	-2.80	-5.07
7.2+4.9sox10:GFP array results	*t*-statistic	-8.8155965805	-7.1221790314
	p-value	0.000021572	0.0000997859
	CLEAR result	diff.expr	diff.expr
	Fold Change	-2.34	-3.15

Table 34: Summary of *aldh2a/b* information

The Affymetrix probe set Dr.4751.1.S1_a_at corresponds to transcripts from two related genes, *aldh2a* and *aldh2b*. The probe set Dr.4751.2.A1_at corresponds to *aldh2b* only. Both probe sets were detected as differentially regulated in all three data sets by both statistical tests (**Table 34**). It seems likely that the differentially expressed gene being detected was *aldh2b* as both probe sets correspond to this gene; it cannot be excluded that *aldh2a* is also differentially regulated. The *in situ* probe used to validate these probe sets was designed to recognise *aldh2b* mRNA (**Figure 35**). At both 24 hpf and 30 hpf, numerous cells expressing *aldh2b* were scattered over the head and anterior trunk of WT embryos (**Figure 35A** and **D**) and cells were visible migrating on the lateral pathway in the trunk at both stages (arrows in **Figure 35A**, **B**, **D** and **E**). No cells migrating on the medial pathway were seen at 24 or 30 hpf (data not shown and **Figure 35F**). A subset of pre-migratory NCCs expressing *aldh2b* was also visible in the trunk at both time points in WT embryos (**Figure 35B**, **C** and **F**). *Aldh2b* expression did not overlap with melanocytes. *Aldh2b* expression was mostly absent in *sox10*$^{-/-}$ embryos (**Figure 35A'**, **C'**, **D'**, **E'** and **F'**)

although some cells in pre-migratory positions were weakly marked (**Figure 35B'**). Thus this gene appears to be expressed in early cells of the xanthophore lineage. The expression pattern for *aldh2b* shown here mostly matched the pattern published on the ZFIN website; a*ldh2b* was noted as being expressed in the NC and in cells on the lateral pathway but in addition weak expression was noted in the otic vesicle (Thisse et al., 2004). *Aldh2a* expression in developing zebrafish embryos has been examined by RT-PCR but NC expression was not assessed (Song et al., 2006). In summary, *aldh2b* is expressed in xanthoblasts and is strongly reduced in *sox10* mutant embryos.

In humans, ALDH2 is responsible for metabolising aldehydes such as acetaldehyde, the oxidative metabolite of alcohol. This intermediate in the breakdown of alcohol is more toxic than alcohol and is thus responsible for many of the symptoms of a hangover. Zebrafish have two orthologues (*aldh2a* and *aldh2b*) of the *ALDH2* gene, a member of the aldehyde dehydrogenase superfamily (Lassen et al., 2005, Song et al., 2006). No link between *ALDH2* and NC biology has been established. Therefore the function of *aldh2b* in zebrafish NC is of great interest. There appears to be no immediately obvious link between alcohol metabolism and NC biology therefore this is a surprising result.

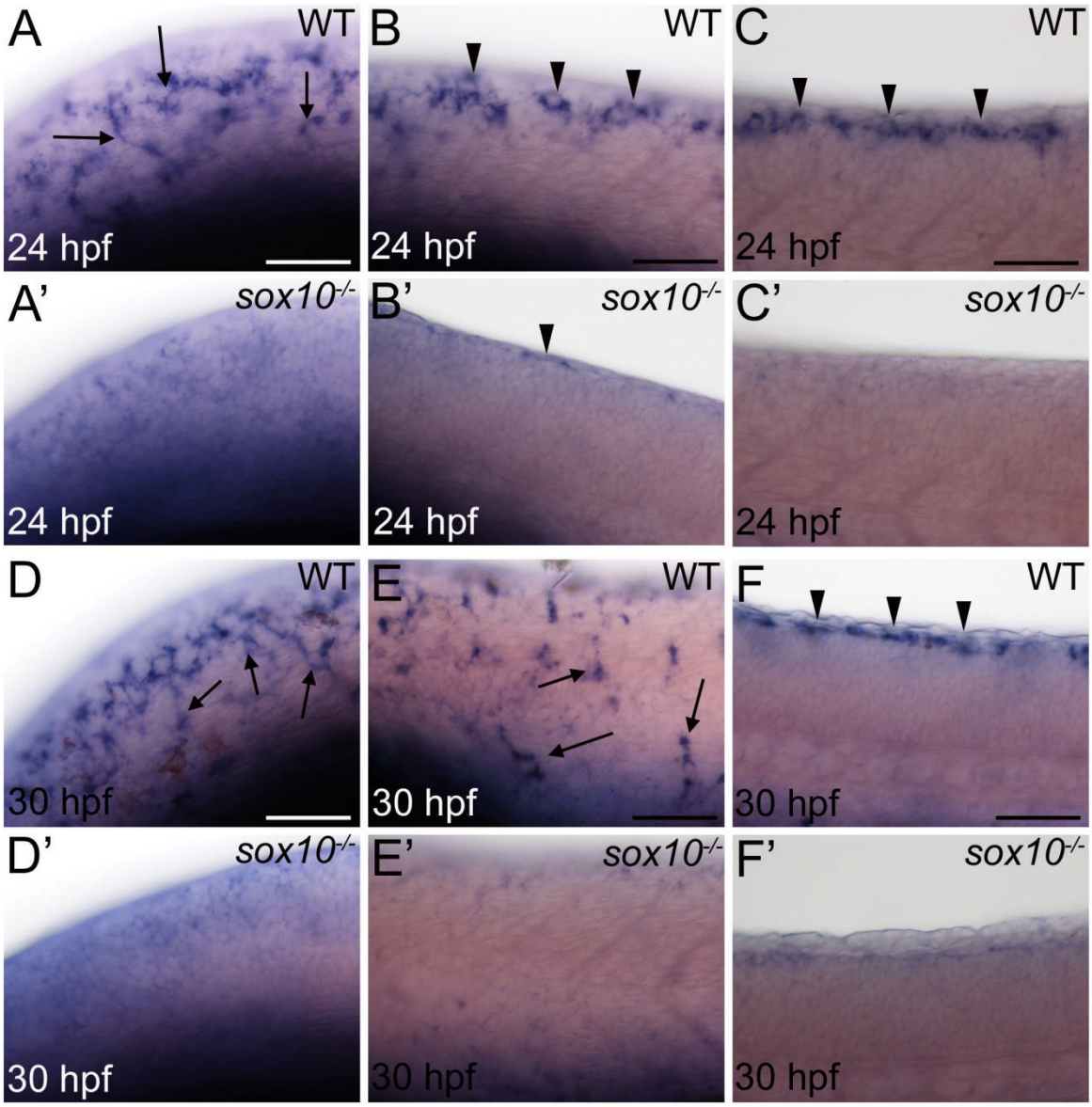


Figure 35: Whole mount *in situ* hybridization pattern of *aldh2b* expression.
All images are presented as per Figure 20. Images A, A', D and D' are of posterior head and anterior trunk showing *aldh2b* expressing cells scattered over this region in WT embryos and migrating on the lateral pathway (arrows). Images B, B', E and E' are of anterior trunk, at 24 hpf and 30 hpf pre-migratory NCCs expressing *aldh2b* were visible (arrowheads) and at 30 hpf cells were migrating on the lateral pathway (arrows). Images C, C', F, and F' are of mid-trunk showing pre-migratory NCCs expressing *aldh2b* at both 24 and 30 hpf (arrowheads). No cells expressing *aldh2b* were visible on the medial pathway (F). Expression of *aldh2b* was mostly absent in *sox10* mutant embryos at both time points (A', B', C', D', E' and F'). All scale bars = 50 µm.

4.2.5.4 Otic Vesicle Expressed Genes

This section contains genes that were expressed in the otic vesicle of zebrafish embryos. It was predicted that these genes would be identified as differentially regulated in the *7.2sox10*:GFP line data but not the *4.9sox10*:GFP data. This is because the otic vesicle is only weakly marked by GFP in the *4.9sox10*:GFP transgenic line (**Figure 7**).

4.2.5.4.1 *cldnj*

Affymetrix Probe Identifier		Dr.8680.1.S1_at
Affymetrix Gene Symbol		*cldnj*
ZFIN Gene Symbol		*cldnj*
Full Name		*claudin j*
Orthology		N/A
Biological Process GO Term		lifelong otolith mineralization
Molecular Function GO Term		structural molecule activity
7.2sox10:GFP array results	*t*-statistic	-16.1235923767
	p-value	0.0000865462
	CLEAR result	diff.expr
	Fold Change	-2.71
4.9sox10:GFP array results	*t*-statistic	-2.44237113
	p-value	0.1346058547
	CLEAR result	diff.expr
	Fold Change	-1.51
7.2+4.9sox10:GFP array results	*t*-statistic	-2.0496237278
	p-value	0.0745517164
	CLEAR result	diff.expr.high.var
	Fold Change	-2.35

Table 35: Summary of *cldnj* information.

The Affymetrix probe set Dr.8680.1.S1_at corresponds to the gene *cldnj*. The *t*-Test declared this gene as differentially regulated in the 7.2 data set only; the CLEAR test indicated that this gene is differentially regulated in all three data sets. The gene expression pattern of *cldnj* was determined by *in situ* hybridization (**Figure 36**). *Cldnj* was strongly expressed in the otic epithelium of WT 24 hpf and 30 hpf embryos (**Figure 36A** and **B**). Expression in the otic epithelium was not uniform; expression in the anterior of the otic vesicle was reduced in comparison with the posterior. This expression pattern has been previously characterised and matches the pattern shown here (Hardison et al., 2005). *Cldnj* expression was strongly reduced in 24 hpf and 30 hpf *sox10*$^{-/-}$ embryos (**Figure 36A'** and **B'**). The distribution of *cldnj* expression was thoroughly examined by Hardison et al., 2005, they concluded that *cldnj* was expressed in the dorso-lateral areas of the otic epithelium. Thus *cldnj* is strongly expressed in the otic epithelium and for the first time has been shown to be down-regulated in *sox10* mutant embryos. It is likely the down-regulation of *cldnj* in *sox10* mutant embryos contributes to the small otoliths phenotype of these embryos.

The zebrafish *cldnj* mutant was generated from a large scale insertional mutant screen and subsequently characterised (Amsterdam et al., 2004, Hardison et al., 2005). The *cldnj* mutant had reduced otoliths apparently as a result of failing to properly incorporate otolith material. The mechanism by which *cldnj* functions was not elucidated. The *cldnj* mutant displays reduced hearing and vestibular function (balance and spatial orientation). In vertebrates, otoliths respond to changes in the direction of gravity as a result of head movements. This distorts the hairs on sensory cells in the macula to detect the movement. It is therefore not surprising that *cldnj* mutants display vestibular problems. In lower vertebrates the otoliths also function to detect sound thus the zebrafish hearing defect fits with a defect in otoliths formation (Finn and Kapoor, 2008). *Sox10*$^{-/-}$ embryos display a small otic vesicle and small otolith phenotype (Whitfield et al., 1996) and thus the *cldnj* mutant displays an element of the *sox10* mutant phenotype. It would be interesting to assess how *sox10* mutant embryos perform in the tapping (hearing) test and startle response test used to assess *cldnj* mutants. There is no clear human orthologue for the zebrafish *cldnj* gene (Hardison et al., 2005). However mutations in the gene *CLDN14* result in human deafness (Wilcox et al., 2001) thus there is a precedent for this gene family to function in human hearing. In summary, *cldnj* is a relatively well characterised gene known to be expressed in the otic epithelium and the phenotype of a *cldnj* mutant has been characterised.

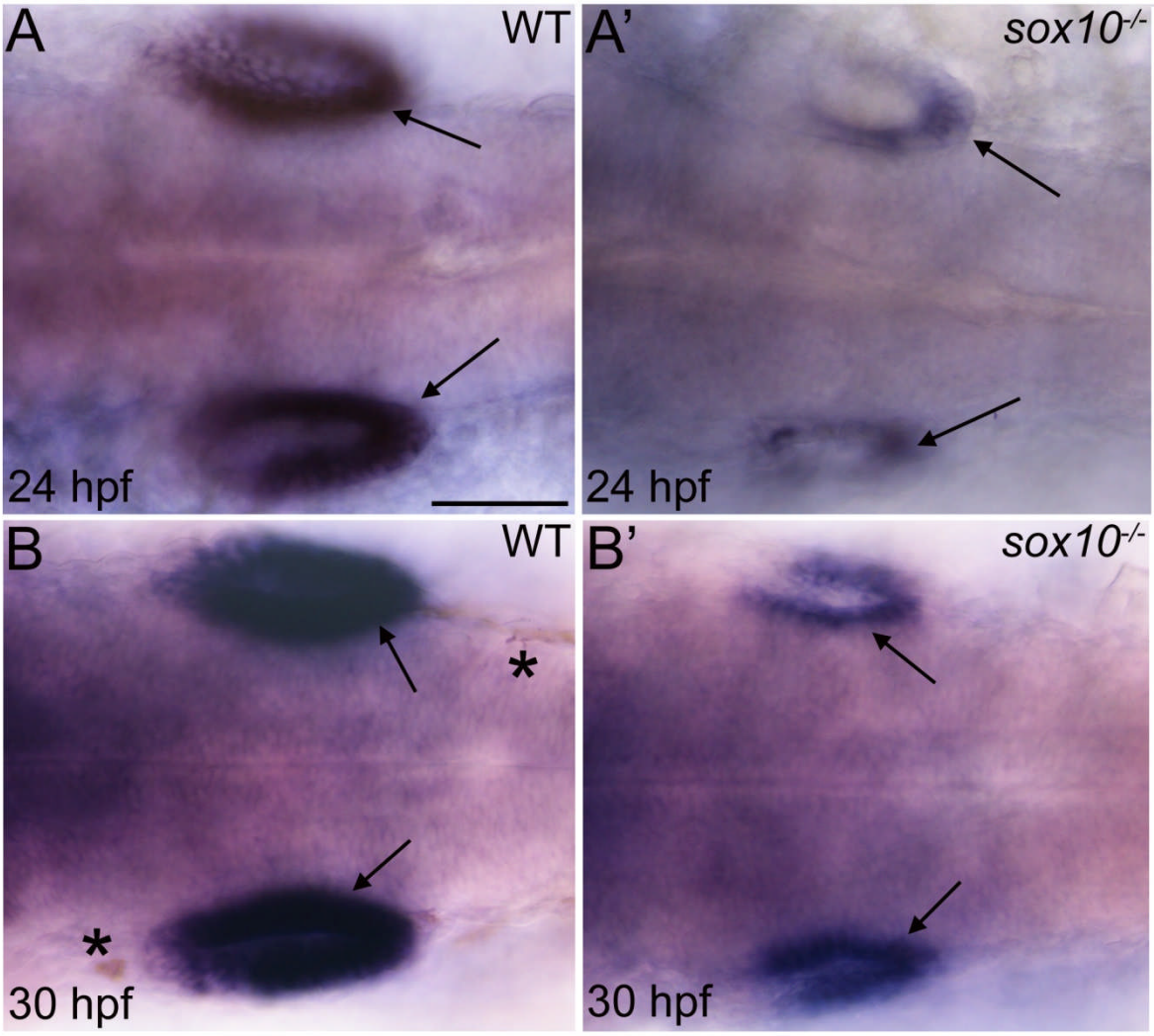


Figure 36: Whole mount *in situ* hybridization pattern of *cldnj* expression.
All images are dorsal views of embryos oriented anterior left and posterior right. Images A and A' are of 24 hpf embryos and images B and B' are of 30 hpf embryos. Images A and B are of WT embryos, images A' and B' are of *sox10* mutant embryos. At both 24 hpf and 30 hpf the otic epithelium strongly expressed *cldnj* (arrows in A and B), the presence of melanocytes is indicated (*). At these time points *cldnj* expression in the otic epithelium was strongly reduced in *sox10* mutants (arrows in A' and B'). Scale bar = 75 µm.

4.2.5.4.2 *otomp*

Affymetrix Probe Identifier		Dr.19416.1.S1_at
Affymetrix Gene Symbol		*otomp*
ZFIN Gene Symbol		*otomp*
Full Name		*otolith matrix protein*
Orthology		N/A
Biological Process GO Term		iron ion transport cellular iron ion homeostasis negative regulation of signal transduction otolith development otolith development
Molecular Function GO Term		ferric iron binding
7.2sox10:GFP array results	*t*-statistic	-6.6376228333
	p-value	0.0026735307
	CLEAR result	diff.expr
	Fold Change	-1.86
4.9sox10:GFP array results	*t*-statistic	-9.0619659424
	p-value	0.0119594019
	CLEAR result	diff.expr
	Fold Change	-1.50
7.2+4.9sox10:GFP array results	*t*-statistic	-3.0389389992
	p-value	0.0160892177
	CLEAR result	diff.expr
	Fold Change	-1.96

Table 36: Summary of information for *otomp*.

The Affymetrix probe set Dr.19416.1.S1_at corresponds to the gene *otomp*. *Otomp* is a strongly expressed marker of the otic epithelium (Thisse et al., 2004, Murayama et al., 2005) and is down-regulated in *sox10*$^{-/-}$ embryos (R. Kelsh, personal communication). Thus this gene has served as the only ear specific marker gene throughout this project. *Otomp* was identified as differentially expressed in all three data sets by both statistical tests (**Table 36**). Validation of *otomp* down-regulation in *sox10* mutant embryos was performed by *in situ* hybridization (**Figure 37**). As previously reported, *otomp* was specifically expressed throughout the otic epithelium with expression more intense in a ventro-posterior area (**Figure 37A**, **B**, **C** and **D**) (Murayama et al., 2005). This was true at both 24 hpf and 30 hpf. In *sox10*$^{-/-}$ embryos *otomp* expression persisted at both 24 and 30 hpf but in a more restricted ventro-posterior region (**Figure 37A'**, **B'**, **C'** and **D'**). Thus *otomp* was expressed in the otic epithelium and was down-regulated in *sox10* mutant embryos.

The function of *otomp* in zebrafish has been studied by morpholino mediated gene knockdown (Murayama et al., 2005). *Otomp* morphants displayed otoliths that were reduced in size compared to control embryos and also displayed behavioural defects consistent with impaired vestibular functions. Otoliths in morphant embryos initially formed correctly (seeded) but grew at a reduced rate in comparison with WT embryos. This otolith

phenotype therefore shares similarities with the *sox10* mutant small otoliths phenotype and the reduction of *otomp* expression in these embryos is likely to contribute to the small otolith phenotype.

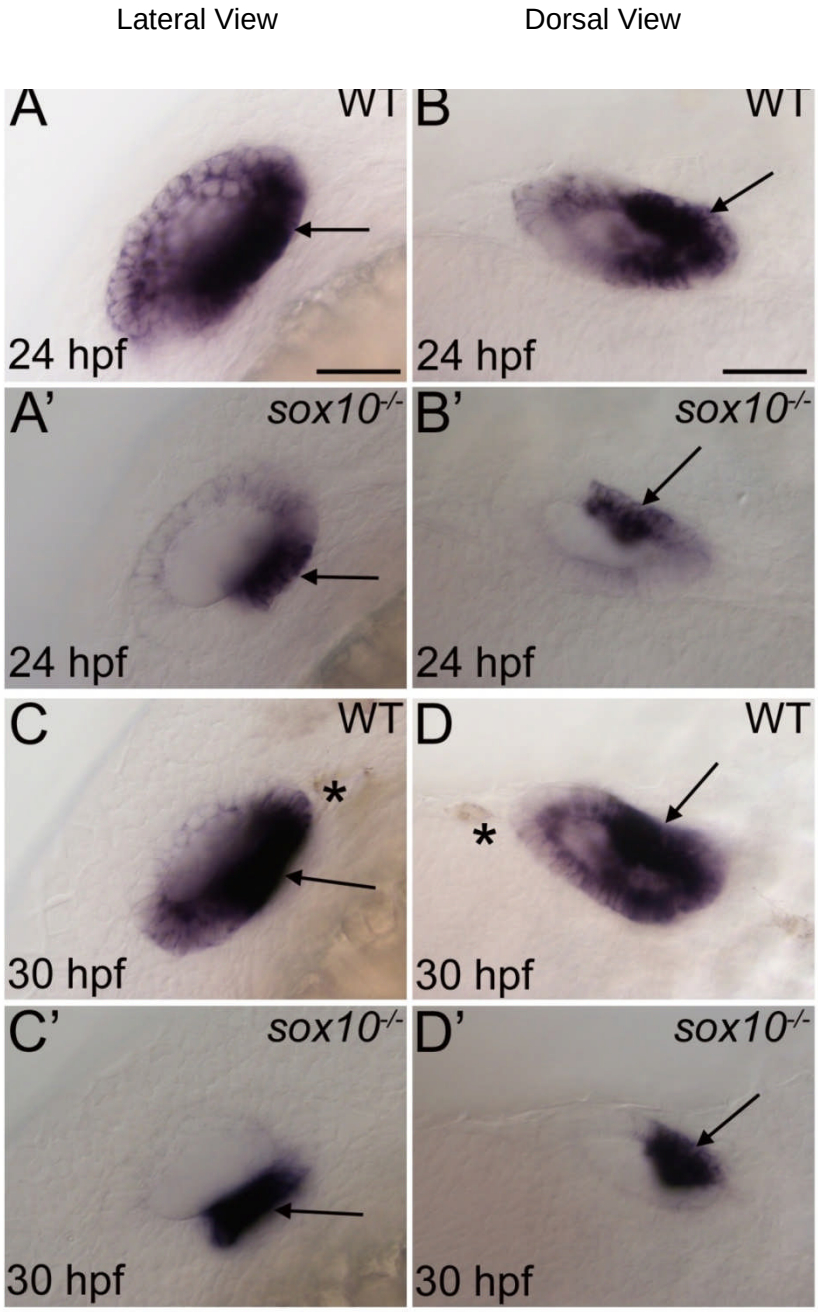


Figure 37: Whole mount *in situ* hybridization pattern of *otomp* expression.
All images are of zebrafish otic vesicles. Images A, A', C and C' are lateral views oriented anterior down and ventro-posterior right. Images B, B', D and D' are dorsal views oriented anterior left. In WT embryos *otomp* was expressed throughout the otic epithelium but was strongly expressed in a ventro-posterior region (arrows in A, B, C and D). In mutant embryos expression was limited to a more restricted ventro-posterior region (arrows in A', B', C' and D'). Melanocytes in WT 30 hpf embryos are indicated (*). Scale bars = 60 µm.

4.2.5.4.3 Dr.18158

Affymetrix Probe Identifier		Dr.18158.1.A1_at
Affymetrix Gene Symbol		N/A
ZFIN Gene Symbol		N/A
Full Name		N/A
Orthology		N/A
Biological Process GO Term		N/A
Molecular Function GO Term		N/A
7.2sox10:GFP array results	*t*-statistic	-7.7104620934
	p-value	0.0015227671
	CLEAR result	diff.expr
	Fold Change	-1.53
4.9sox10:GFP array results	*t*-statistic	-2.5671079159
	p-value	0.1241162792
	CLEAR result	diff.expr
	Fold Change	-1.45
7.2+4.9sox10:GFP array results	*t*-statistic	-1.4548796415
	p-value	0.1837879419
	CLEAR result	diff.expr.high.var
	Fold Change	-1.73

Table 37:Summary of Results for Dr.18158.1.A1_at.

The Affymetrix probe set Dr.18158.1.A1_at has no associated annotation on NetAffx. In a similar pattern of results to *cldnj*, the CLEAR test identified this gene as differentially regulated in all three data sets while the *t*-Test only identified this gene as differentially regulated in the 7.2 data set (**Table 37**). To validate these results an *in situ* hybridization probe was designed using the Dr.18158.1.A1_at sequence on NetAffx (**Figure 38**). Dr.18158 was expressed mainly in a ventro-medial region of the otic epithelium in 24 and 30 hpf WT embryos but expression was also observed in a ventro-lateral region in the posterior of the otic vesicle (**Figure 38A**, **B**, **C** and **D**). In *sox10* mutant embryos the same expression pattern was observed but expression was noticeably reduced (**Figure 38A'**, **B'**, **C'** and **D'**). Thus the gene that Dr.18158.1.A1_at corresponds to is expressed in the otic epithelium and is down-regulated in *sox10* mutant embryos.

The Dr.18158.1.A1_at probe set has not been annotated so no information regarding the gene that this probe set corresponds to is available. Performing BLAST searches on the NCBI and ZFIN websites does not show any strong similarity between the Affymetrix sequence for the probe set Dr.18158.1.A1_at and any mRNA or cDNA sequences. The sequence of Dr.18158.1.A1_at is contained within the GenBank genomic DNA sequence BX530036. As such this is a completely novel target to have been detected and validated as part of this microarray project. Functional knock down for this gene using a morpholino approach to uncover the role of this gene would be very interesting. Such an approach could help to unravel another aspect of the complex ear phenotype of *sox10* mutant embryos.

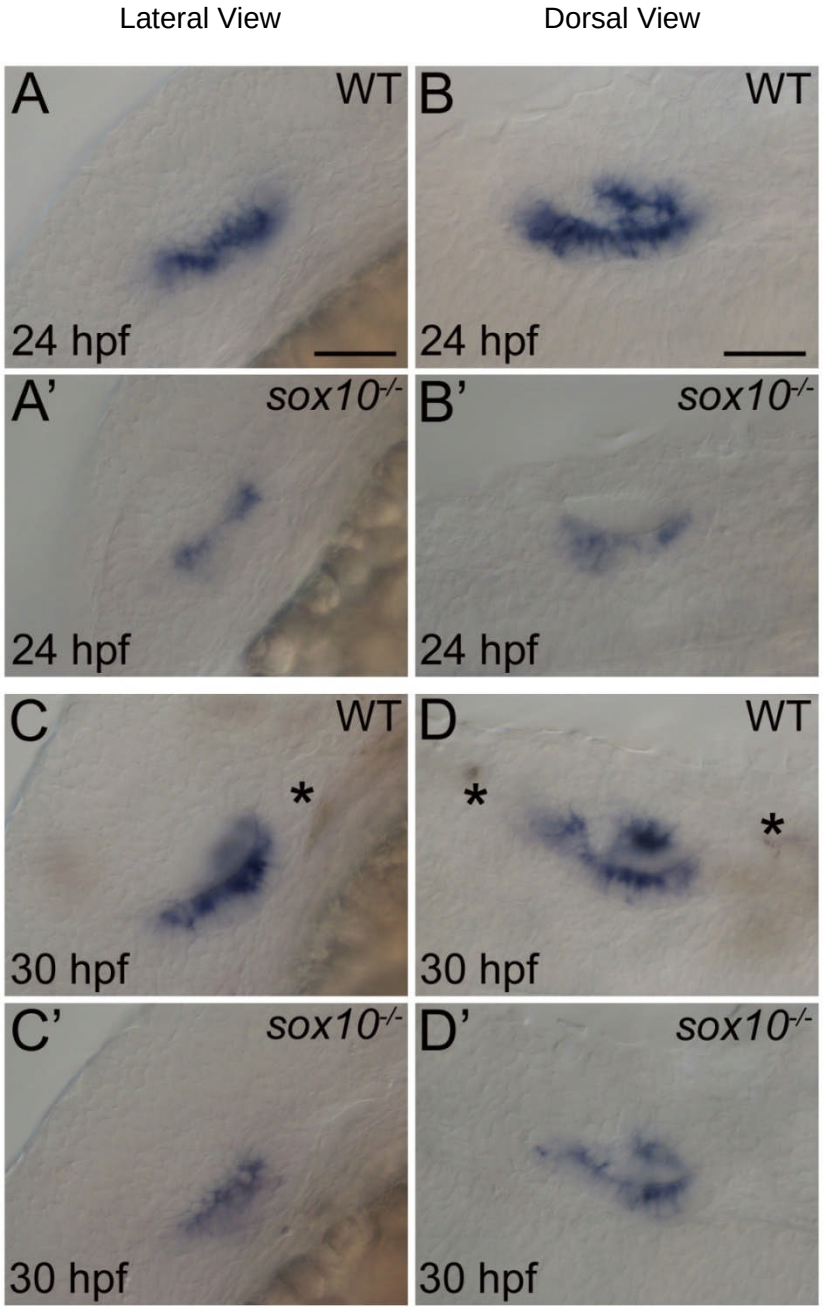


Figure 38: Whole mount *in situ* hybridization pattern of Dr.18158.1.A1_at expression.
All images are of zebrafish otic vesicles presented as per Figure 37. In WT embryos at both 24 hpf and 30 hpf Dr.18158 was expressed mainly in a ventro-medial region but also in a ventro-lateral position in the posterior of the otic vesicle (A, B, C and D). In mutant embryos the same expression pattern was observed but expression was comparatively weaker (A', B', C' and D'). Melanocytes in WT 30 hpf embryos are indicated (*). Scale bars = 60 µm.

4.2.5.5 foxD3 is down-regulated in the peripheral nervous system.

Affymetrix Probe Identifier		Dr.590.1.S1_at
Affymetrix Gene Symbol		*foxd3*
ZFIN Gene Symbol		*foxd3*
Full Name		*forkhead box D3*
Orthology		*FOXD3*
Biological Process GO Term		neural crest cell migration regulation of transcription, DNA-dependent peripheral nervous system development neural crest cell development melanocyte differentiation enteric nervous system development sympathetic nervous system development lateral line nerve glial cell development pigment cell differentiation iridophore differentiation cartilage development
Molecular Function GO Term		DNA binding transcription factor activity transcription activator activity sequence-specific DNA binding
7.2sox10:GFP array results	*t*-statistic	-8.5286149979
	p-value	0.0010371582
	CLEAR result	diff.expr
	Fold Change	-1.32
4.9sox10:GFP array results	*t*-statistic	-3.1133506298
	p-value	0.0895293579
	CLEAR result	diff.expr
	Fold Change	-1.50
7.2+4.9sox10:GFP array results	*t*-statistic	-3.1978538036
	p-value	0.0126530966
	CLEAR result	diff.expr
	Fold Change	-1.32

Table 38: Summary of information for *foxd3*.

The Affymetrix probe set Dr.590.1.S1_at corresponds to the transcription factor *foxd3*. This gene was determined as differentially regulated by both statistical tests on all data sets except the *t*-Test on the 4.9 data set (**Table 38**). To determine if *foxd3* was differentially regulated in *sox10* mutant embryos, *in situ* hybridization was performed (**Figure 39**). *Foxd3* is a well known early NC marker and was expressed in NCCs in the posterior trunk of WT 24 hpf embryos (**Figure 39D**) (Kelsh et al., 2000, Odenthal and Nüsslein-Volhard, 1998). Expression in early pre-migratory NCCs appeared to be identical in WT and *sox10*$^{-/-}$ embryos at 24 hpf and 30 hpf (**Figure 39D**, **D'** and data not shown) (Lopes et al., 2008). *Foxd3* was also expressed in clumps of cells anterior of the otic vesicle (**Figure 39A**), posterior of the otic vesicle (**Figure 39B**) and in cells along the posterior lateral line nerve (**Figure 39B**). These sites showed a reduction in *foxd3* expression in *sox10* mutant embryos (**Figure 39A'** and **B'**). Previous studies have also identified that *foxd3* is down-regulated at these sites in *sox10* mutant embryos and

postulated that these cells are of the glia lineage (Kelsh et al., 2000, Lister et al., 2006). F*oxd3* expression was also noted in cells migrating on the medial pathway (**Figure 39C**). These cells are in a position consistent with dorsal root ganglia (DRG) and are likely again to be of the glia lineage (Kelsh et al., 2000). Obvious down-regulation of *foxd3* expression in DRG associated cells was not detected in *sox10*$^{-/-}$ embryos (**Figure 39C'**). *FoxD3* expression in somites was also observed; this site displayed no difference in expression between WT and mutant embryos (**Figure 39D** and **D'**). It has been shown that *foxd3* is expressed in somites and regulates the muscle differentiation gene *myf5* during somitogenesis (Lee et al., 2006). The expression pattern of *foxd3* in 30 hpf WT and mutant embryos had strong similarities with the expression patterns identified at 24 hpf (data not shown). Thus the expression pattern of *foxd3* in early NCCs and NC derived glia of the peripheral nervous system shown here (**Figure 39**) matches previously published reports. Down-regulation of *foxd3* expression in *sox10* mutant embryos was observed in glia but not early NCCs. The identification of a gene expressed in the PNS as down-regulated in this screen shows that it is possible to detect PNS genes by this approach at this time point.

Defects in *foxd3* expression in *sox10* mutant embryos appear to be restricted to the glial cell lineage of the peripheral nervous system (**Figure 39**) (Kelsh et al., 2000). Glial cell differentiation fails in *sox10* mutant embryos (Carney et al., 2006) thus a reduction in *foxd3* positive glial cells was expected. The number of peripheral glia is also reduced in *foxd3* mutant embryos (Montero-Balaguer et al., 2006, Stewart et al., 2006). A sub-population of NCCs in the hindbrain region die by apoptosis in the *foxd3*^{sym1} mutant (Stewart et al., 2006). The position of these cells is consistent with a glial, neuronal and craniofacial cartilage fate. Thus *foxd3* may play a role in the survival of glia but this requires further investigation. In addition it remains to be seen if this role can be generally extended to trunk NC derived glia associated with the posterior lateral line nerve and the DRGs. It is interesting to note that clear down-regulation of *foxd3* was observed in cells associated with the posterior lateral line nerve but not in cells associated with the DRGs. As *sox10* functions in glial differentiation but not glial specification it is possible that expression of *foxd3* is only defective in differentiated glia and not glia precursor cells. Thus glia precursor cells might still be visualised using *foxd3* as a marker. These DRG associated glia do fail to differentiate in *sox10* mutants by 5 dpf (Carney et al., 2006). Characterisation of the role *foxd3* plays in the development of this NC derivative has been complicated by an early role in pre-migratory NCCs. For example, analysis of *foxd3* mutants identified a down-regulation of important early NC transcription factors including *sox10* (Montero-Balaguer et al., 2006, Stewart et al., 2006). Thus it appears that *foxd3* is expressed and functioning both upstream and downstream of *sox10*. It is likely that these two roles, the role in early NC development and the role in glia are distinct.

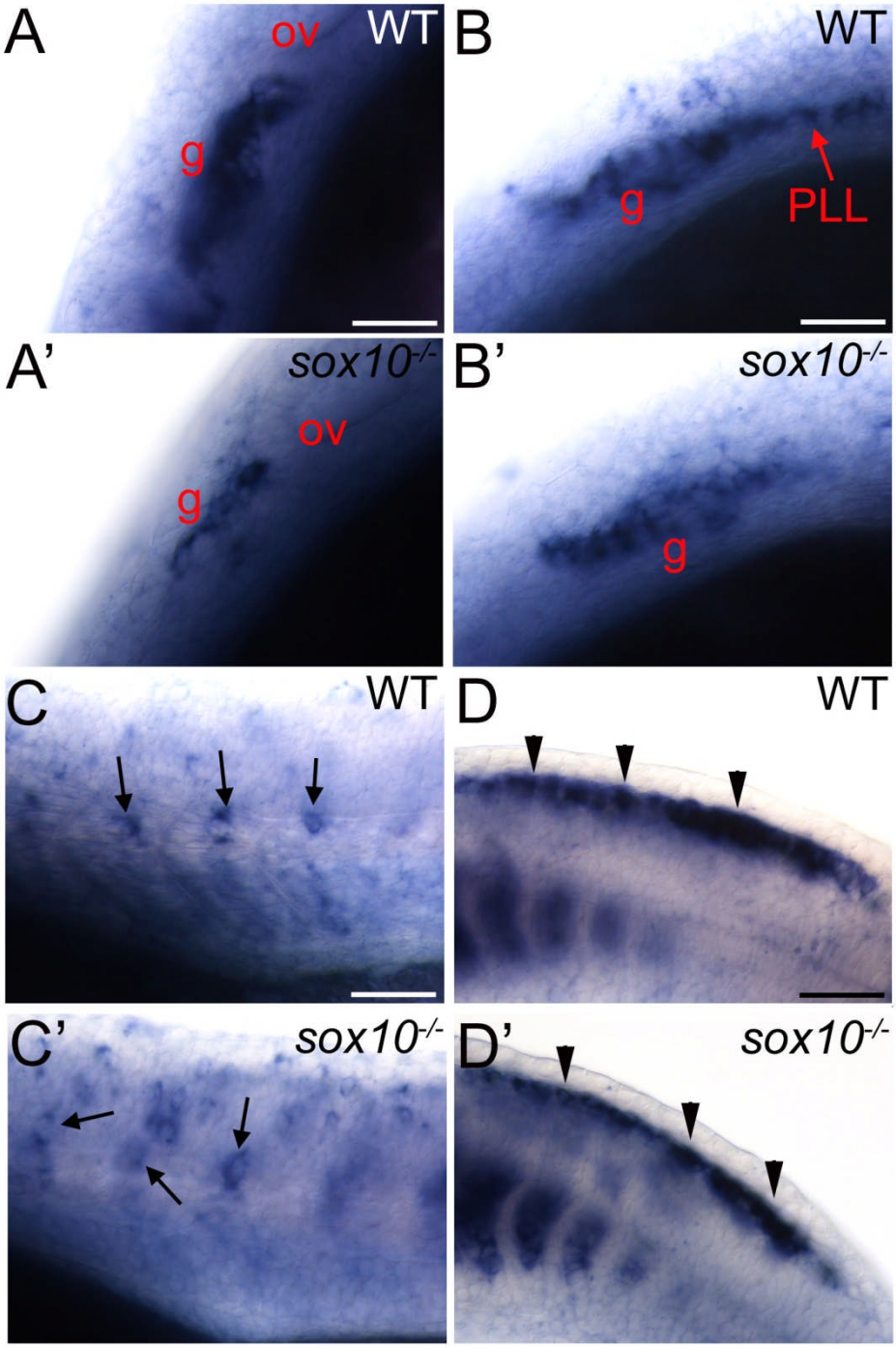


Figure 39: Whole mount *in situ* hybridization pattern of *foxD3* expression in NC derivatives.
All images are lateral views of 24 hpf embryos oriented anterior left and dorsal top except A and A' oriented anterior bottom and dorsal left. Image A and B are of the head, C of anterior mid-trunk and D of posterior trunk (WT embryos), images labelled by a letter with an apostrophe are corresponding images of *sox10* mutant embryos. In WT embryos *foxD3* was expressed in cells anterior of the otic vesicle (A), posterior of the otic vesicle (B) and surrounding the posterior lateral line nerve (B). Expression at these sites was reduced in mutant embryos (A' and B'). In the trunk of WT embryos putative dorsal root ganglion cells were visible in both WT and mutant embryos (arrows in C and C'). *FoxD3* was strongly expressed in early pre-migratory crest in both WT and *sox10* mutant embryos in posterior trunk (arrowheads in D and D'). Labels: g = glia, PLL = posterior lateral line nerve and ov = otic vesicle. Scale bars = 50 µm.

4.2.5.6 Miscellaneous validated down-regulated genes.

Gene placed in to this section show expression in NCCs that is reduced in *sox10* mutant embryos. For some of these genes, the weakness of the expression pattern obtained has made it difficult to accurately predict which cell type the gene is expressed. For the remaining genes, the gene expression pattern does not neatly fit with the expected expression pattern for one NC cell type.

4.2.5.6.1 Dr.4612

Affymetrix Probe Identifier		Dr.4612.1.A1_at
Affymetrix Gene Symbol		*wu:fc31e04*
ZFIN Gene Symbol		*wu:fc31e04*
Full Name		*wu:fc31e04*
Orthology		N/A
Biological Process GO Term		N/A
Molecular Function GO Term		N/A
7.2sox10:GFP array results	*t*-statistic	-18.1266307831
	p-value	0.0000544656
	CLEAR result	diff.expr
	Fold Change	-2.83
4.9sox10:GFP array results	*t*-statistic	-6.9993152618
	p-value	0.0198077019
	CLEAR result	diff.expr
	Fold Change	-4.61
7.2+4.9sox10:GFP array results	*t*-statistic	-10.1082744598
	p-value	0.000007832
	CLEAR result	diff.expr
	Fold Change	-3.41

Table 39: Summary of information for Dr.4612.1.A1_at.

The Affymetrix probe set Dr.4612.1.A1_at corresponds to the expressed sequence *wu:fc31e04*. This probe set was identified as differentially regulated in all three data sets by both statistical tests (**Table 39**). *In situ* hybridization was performed to visualise the expression pattern of this target in WT and *sox10* mutant embryos (**Figure 40**). Dr.4612 was expressed in cranial NCCs at the midbrain-hindbrain boundary, no expression was detected in mutant embryos (**Figure 40A** and **A'**). In the trunk, expression was detected in pre-migratory NCCs (**Figure 40B**) and in cells migrating on the lateral pathway (**Figure 40C**). In addition, Dr.4612 expression was detected in rare cells on the medial pathway, one cell in approximately half of the embryos assessed (**Figure 40C**). No expression was detected at these sites in mutant embryos (**Figure 40B'**, **C'** and **D'**). No expression patterns have been published for this gene. The expression pattern of Dr.4612 (*wu:fc31e04*) has been assessed for the first time here and shown to mark NCCs, particularly cranial NCCs. The expression of this gene was down-regulated in *sox10*$^{-/-}$ embryos.

As a result of the staining being very weak, this pattern proved difficult to interpret. The pattern does show similarities to markers of the xanthophore lineage. For example, most migratory cells expressing Dr.4612 were on the lateral pathway with only rare cells observed on the medial pathway. Other markers of the xanthophore lineage also marked rare medial pathway cells such as *cx33.8* (**Figure 29**). Thus a tentative prediction is that Dr.4612 is expressed in the xanthophore cell lineage. This prediction requires further testing to be validated, for example optimisation of the *in situ* hybridization protocol to improve the signal strength and comparison of Dr.4612 expression in xanthophore and melanocyte pigment mutants. A BLAST search for RNA sequences on the NCBI website using the ZFIN sequence for *wu:fc31e04* did not identify any hits with similarities. Therefore it is impossible to formulate a hypothesis regarding the function of this gene. In summary, although the cell lineage that this gene is expressed in has not been definitively determined, this gene is clearly NC expressed and down-regulated in *sox10* mutant embryos when compared to WT embryos. This target represents a novel NC expressed gene.

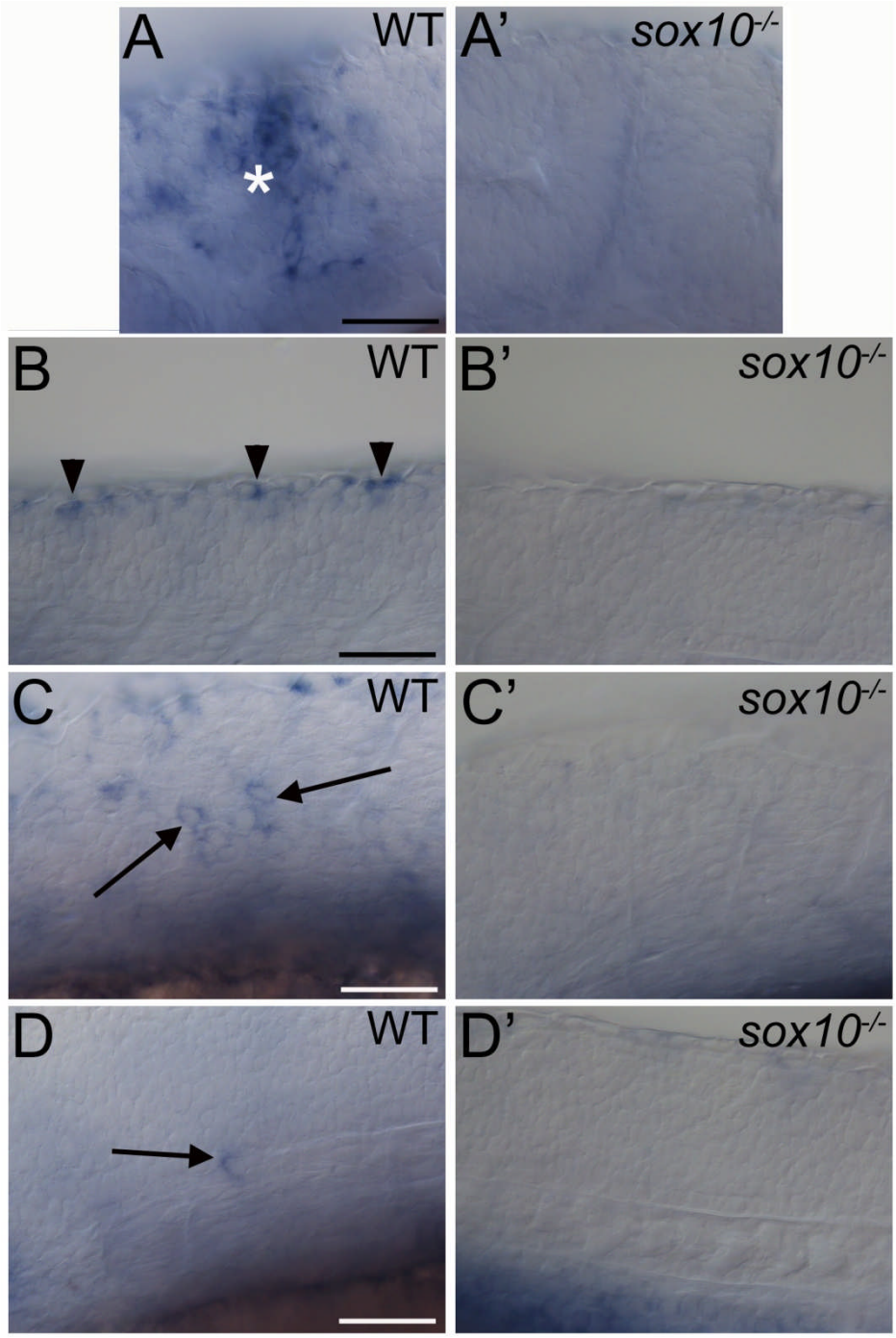


Figure 40: Whole mount *in situ* hybridization pattern for the probe set Dr.4612.1.A1_at.
All images are lateral views of 24 hpf embryos oriented anterior left and dorsal top. Images labelled by a letter alone are of WT embryos; corresponding images of *sox10* mutant embryos are labelled with a letter and apostrophe. Image A and A' are of the head in the region of the midbrain-hindbrain boundary, all other images are of trunk regions. Dr.4612 expressing cells were detected at and migrating near the midbrain hindbrain boundary (asterisk in A). Pre-migratory NCCs expressed Dr.4612 (B) as did migratory cells on the lateral pathway (C) and rare cells on the medial pathway (D). Expression was absent in mutant embryos (A', B', C' and D'). Scale bars = 50 µm.

4.2.5.6.2 *coro1c*

Affymetrix Probe Identifier		Dr.2901.1.S1_at
Affymetrix Gene Symbol		*coro1c*
ZFIN Gene Symbol		*coro1c*
Full Name		*coronin, actin binding protein, 1c*
Orthology		*CORO1C (CORO3)*
Biological Process GO Term		N/A
Molecular Function GO Term		N/A
7.2sox10:GFP array results	*t*-statistic	-9.323184967
	p-value	0.0007367123
	CLEAR result	diff.expr
	Fold Change	-1.45
4.9sox10:GFP array results	*t*-statistic	-1.8859200478
	p-value	0.1999538839
	CLEAR result	non.significant
	Fold Change	-1.15
7.2+4.9sox10:GFP array results	*t*-statistic	-3.5967919827
	p-value	0.0070149088
	CLEAR result	diff.expr
	Fold Change	-1.31

Table 40: Summary of information for *coro1c.*

The Affymetrix probe set Dr.2901.1.S1_at corresponds to the gene *coro1c*, also known as *coronin 3*. This gene was detected as differentially regulated by both statistical tests in the 7.2 and 7.2+4.9 data sets but was not identified as differentially regulated in the 4.9 data set (**Table 40**). To validate differential gene expression in *sox10* mutant embryos, *in situ* hybridization was performed (**Figure 41**). *Coro1c* was expressed in WT 24 hpf embryos in pre-migratory NCC positions (**Figure 41A**) and in cells migrating on the medial pathway (**Figure 41B**). No *coro1c* expression was observed in cells migrating on the lateral pathway (**Figure 41C**). Expression was strongly reduced in pre-migratory NCCs of mutant embryos (**Figure 41A'**) and fewer cells expressing *coro1c* were observed on the medial pathway (**Figure 41B'**). This pattern and location of cells suggested that *coro1c* could be expressed in cells of a neural lineage as these cells only migrate on the medial pathway and, unlike pigment cells which fail to migrate in *sox10* mutant embryos, these cell types are only reduced in number. At 30 hpf a different pattern of gene expression was identified. *Coro1c* expression was detected in cells migrating on the medial (**Figure 41D** and **E**) and lateral (**Figure 41F**) pathways. Expression of a gene on the lateral pathway is strongly indicative of a pigment cell fate. Often *coro1c* expression overlapped with melanising cells (arrows in **Figure 41D** and **E**) although not always (arrow and asterisk in **Figure 41E**). Expression in migrating cells was absent in mutant embryos (**Figure 41D'**). The expression pattern of *coro1c* at 30 hpf is more consistent with a melanocyte pattern, albeit a restricted pattern when compared to more commonly used markers like *dct*. At both time points the expression of *coro1c* was difficult to interpret as the *in situ* hybridization stain was very weak, especially in the trunk and only a few cells were

labelled. For example, expression of *coro1c* may not be absent in lateral pathway migrating cells in 24 hpf embryos but below obviously detectable levels. In addition to this, strong staining in the head of the CNS was observed (data not shown). This made it difficult to assess for NC *coro1c* expression in the head. If *coro1c* is expressed in cells of a glial lineage then expression would be expected in patches anterior and posterior of the otic vesicle as observed with *foxd3* (**Figure 39**). If *coro1c* is expressed in peripheral neurons the gene expression pattern would be expected to reflect *neurogenin 1* expression. Examination of *coro1c* expression in the *sox10*^{baz1} mutant which has supernumerary DRGs would help to ascertain if this gene is a neuronal marker (Carney et al., 2006). An expression pattern for *coro1c* has been published on the ZFIN website showing strong expression in the CNS particularly the hindbrain but no NC expression was detected (Thisse et al., 2004). In summary, *coro1c* is a novel gene expressed in zebrafish NC that is down-regulated in *sox10* mutant embryos. The cell lineage this gene is expressed in could not be comprehensively identified but it may mark NC derivatives of the PNS or melanocytes. There is a paucity of PNS cell lineage markers therefore it is important to determine if *coro1c* is a PNS marker.

The family of coronin genes are known regulators of the actin cytoskeleton and have been implicated in actin dependent processes including cell migration, reviewed in (Uetrecht and Bear, 2006). *CORO1C* was cloned in 2000 and immunocytochemical staining identified that *CORO1C* co-localised with F-actin indicating possible roles in cytokinesis, cell motility and signal transduction (Iizaka et al., 2000). Functional studies *in vitro* indicated that absence of functional *CORO1C* inhibited cell migration possibly through a participation in the formation of cellular protrusions (Rosentreter et al., 2007). Expression of *CORO1C* was also correlated with brain tumour malignancy, highly malignant tumours contained more *CORO1C* positive cells (Thal et al., 2008). In addition, *CORO1C* knockdown reduced proliferation, cell motility and invasion of cancer cells *in vitro* (Thal et al., 2008). Thus *CORO1C* has been implicated in several processes that are common to NCCs and cancerous cells. Zebrafish have the potential to provide an *in vivo* model for *coro1c* functional studies and therefore to complement the *in vitro* studies performed so far. No link between *coro1c* and NC biology other than demonstrated here exists.

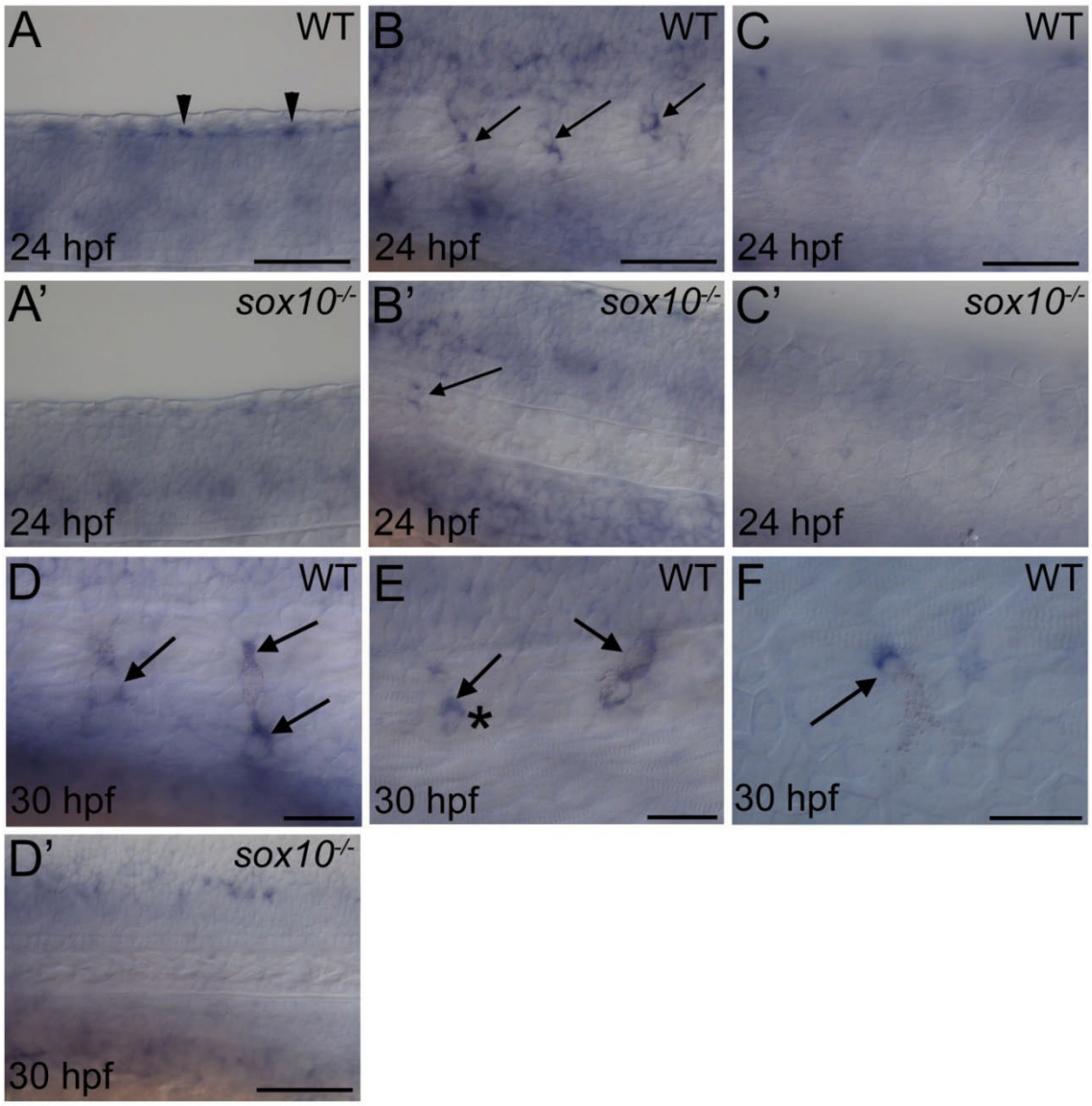


Figure 41: Whole mount *in situ* hybridization pattern of *coro1c* expression.
All images are lateral views of embryonic trunk presented as per Figure 20. Expression of *coro1c* at 24 hpf in WT embryos was detected in pre-migratory NCCs (arrowheads A), NCCs migrating on the medial pathway (arrows in B) but not in cells migrating on the lateral pathway (C). In *sox10* mutant embryos, expression was strongly reduced in pre-migratory NCCs (A') however rare *coro1c* positive cells were detected on the medial pathway (arrow in B'). In 30 hpf WT embryos *coro1c* expression was detected in cells migrating on the medial and lateral pathways (arrows in D, E and F). Melanising cells expressing *coro1c* are indicated with an arrow, a cell expressing *coro1c* but not melanin is marked with an arrow and an asterisk (E). Expression in migratory cells was absent at 30 hpf in *sox10*$^{-/-}$ embryos (D'). Scale bars in D-F = 25 µm, all other scale bars = 50 µm.

4.2.5.6.3 Dr.3972

Affymetrix Probe Identifier		Dr.3972.1.S1_at
Affymetrix Gene Symbol		N/A
ZFIN Gene Symbol		*si:dkey-226k3.4*
Full Name		*si:dkey-226k3.4*
Orthology		N/A
Biological Process GO Term		N/A
Molecular Function GO Term		N/A
7.2sox10:GFP array results	*t*-statistic	-8.3018627167
	p-value	0.0011496505
	CLEAR result	diff.expr
	Fold Change	-3.20
4.9sox10:GFP array results	*t*-statistic	-10.5303659439
	p-value	0.0088978764
	CLEAR result	diff.expr
	Fold Change	-3.25
7.2+4.9sox10:GFP array results	*t*-statistic	-12.9263210297
	p-value	0.0000012141
	CLEAR result	diff.expr
	Fold Change	-3.36

Table 41: Summary information for the Affymetrix probe set Dr.3972.1.S1_at.

The probe set Dr.3972.1.S1_at has not been annotated with a gene name by Affymetrix, a BLAST search of the probe target sequence on ZFIN identified very strong homology (99%) to *si:dkey-226k3.4*. This gene was identified as differentially regulated in all three data sets by both statistical tests (**Table 41**). To validate this, *in situ* hybridization was performed (**Figure 42**). Expression of Dr.3972.1.S1_at was strongly detected in cranial NCCs at the midbrain-hindbrain boundary (**Figure 42A**) and in cells in a NC position in the posterior head and anterior trunk region (**Figure 42B**). Weak expression of Dr.3972.1.S1_at was seen in pre-migratory trunk NCCs and in some cells migrating on the medial pathway (**Figure 42C** and **D**). No expression was detected in cells migrating on the lateral pathway (data not shown). Expression at all these sites was absent in *sox10* mutant embryos (**Figure 42A'**, **B'**, **C'** and **D'**). Expression of Dr.3972.1.S1_at was detected in WT 30 hpf embryos in some cells migrating on the medial pathway (**Figure 42E**) but not on the lateral pathway (data not shown). Occasionally Dr.3972.1.S1_at expression overlapped with melanising cells but this was not the case for the majority of cells. As seen at 24 hpf, no Dr.3972.1.S1_at expression was detected in 30 hpf mutant embryos (**Figure 42E'**).

This expression pattern is hard to interpret as only a few migrating NCCs weakly expressed Dr.3972.1.S1_at. The pattern of Dr.3972.1.S1_at expression in the head had similarities to a marker of pigment cells. Dr.3972.1.S1_at expressing cells were only observed to migrate on the medial pathway. Both cells of the iridophores lineage and neural lineage cells migrate solely on the medial pathway. Cells expressing

Dr.3972.1.S1_at were observed in positions consistent with that occupied by DRGs (Carney et al., 2006). Neural cell fates are only reduced and still migrate in *sox10* mutants while pigment cell fates fail to migrate in mutant embryos. Thus an absence of expression of Dr.3972.1.S1_at in mutants argues that Dr.3972.1.S1_at marks iridophores. The other cell type that migrates on the medial pathway is melanocytes, these also migrate on the lateral pathway, which was not observed for Dr.3972.1.S1_at expressing cells, and Dr.3972.1.S1_at did not typically mark melanising cells. Thus the expression pattern is not consistent with the melanocyte lineage. Dr.3972.1.S1_at expression does mark NCCs, particularly cranial NCCs, is down-regulated in *sox10* mutant embryos. Most likely, this gene marks cells of the iridophore lineage but it cannot be excluded that this gene marks neural cells. Both of these lineages have a dearth of markers and thus it could prove useful to optimise the *in situ* hybridization protocol for this gene and to further characterise the expression pattern obtained here. To test which cell lineage is marked, *in situ* hybridization could be performed in the *ltk*/*shady* mutant that lacks iridophores or in the *sox10*^{baz1} mutant, which has additional sensory neurons. Alternatively to assess for expression in neural cells *in situ* hybridization could be performed in *ngn1* morphant embryos.

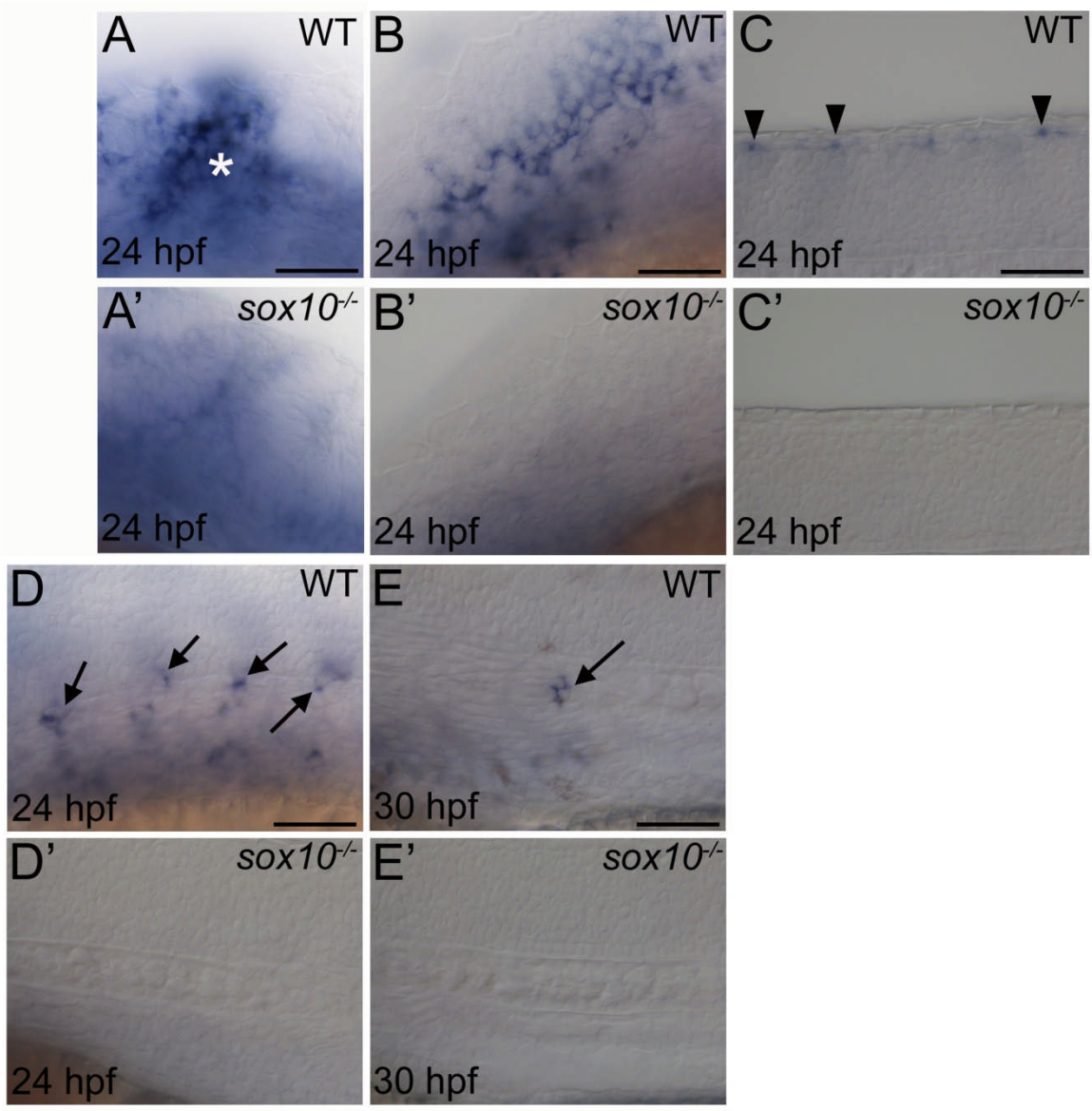


Figure 42: Whole mount *in situ* hybridization visualised expression pattern for Affymetrix probe set Dr.3972.1.S1_at.
All images are lateral views of embryos presented as per Figure 20. Images A and A' are of the head at the midbrain-hindbrain boundary, images B and B' are of posterior head and anterior trunk, all remaining images are of trunk regions. At 24 hpf in WT embryos, expression of Dr.3972.1.S1_at was detected in NCC positions at the midbrain-hindbrain boundary (* in A), in the posterior head (B) and in pre-migratory NCCs in the trunk (arrowheads in C). Expression was also observed in cells migrating on the medial pathway (arrows in D). Expression of Dr.3972.1.S1_at was absent in mutant embryos. In 30 hpf WT embryos, Dr.3972.1.S1_at expressing cells were observed migrating on the medial pathway (arrow in E). This expression was absent in mutant embryos (E'). All scale bars = 50 µm.

4.2.6 Validated Genes Up-Regulated in *sox10* mutant embryos.

4.2.6.1 stc2

Affymetrix Probe Identifier		Dr.23461.1.A1_at
Affymetrix Gene Symbol		*stc2*
ZFIN Gene Symbol		*stc2*
Full Name		*stanniocalcin 2*
Orthology		*STC2*
Biological Process GO Term		N/A
Molecular Function GO Term		hormone activity
7.2sox10:GFP array results	*t*-statistic	7.0355191231
	p-value	0.0021509288
	CLEAR result	diff.expr
	Fold Change	1.53
4.9sox10:GFP array results	*t*-statistic	1.5754812956
	p-value	0.2558328807
	CLEAR result	diff.expr.high.var
	Fold Change	1.66
7.2+4.9sox10:GFP array results	*t*-statistic	3.9415335655
	p-value	0.0042866552
	CLEAR result	diff.expr
	Fold Change	1.59

Table 42: Summary of information for *stc2*.

The Affymetrix probe set Dr.23461.1.A1_at corresponds to the gene *stc2*. This gene was identified as differentially regulated in all three data sets by both statistical tests except the *t*-Test on the 4.9 data set (**Table 42**). To validate that *stc2* was up-regulated in *sox10* mutant embryos, *in situ* hybridization was performed (**Figure 43**). Expression of *stc2* was identified in two ventro-medial patches in the otic vesicle at both 24 hpf and 30 hpf in WT embryos (arrowheads in **Figure 43A** and **D**). One patch appears to be more anterior while the other patch is more posterior. These patches correspond to the position at which the hair cells of the anterior and posterior sensory maculae develop (Dutton et al., 2009). In *sox10* mutant embryos at 24 hpf the same patches of *stc2* expression were visible (**Figure 43A'** and **D'**) although one of the patches is out of focus in **Figure 43A'**. In addition to this, the domain of *stc2* expression had expanded to include an ectopic ventral region of the otic vesicle in between the two previously described patches (**Figure 43A'**). In 30 hpf mutant embryos expression of *stc2* was observed in the same two patches seen in WT embryos but expression typically appeared to be stronger (**Figure 43D'**). Expression in these two patches was very variable in WT embryos at 30 hpf, ranging from weak to strong expression, while in mutant embryos *stc2* expression was consistently strong. *Stc2* expression was also observed in rare cells with a pigment cell like morphology in WT 24 hpf embryos including cells migrating on the medial and lateral pathways (arrows in **Figure 43A**, **B** and **C**). At 30 hpf, *stc2* expression overlapped with some migratory melanising cells on the medial (**Figure 43E**) and lateral pathways (data

not shown). The position of these cells was consistent with a melanocyte fate although only a few cells (approximately 3-5) were marked in each embryo. *Stc2* expression was not detected in these cells in *sox10* mutant embryos (**Figure 43B'** and **E'**). An expression pattern for *stc2* has been published on the ZFIN website showing otic vesicle expression but not NCC expression (Thisse et al., 2004). In summary, *stc2* expression was up-regulated in the otic vesicle of *sox10*$^{-/-}$ embryos thus validating the microarray data. Surprisingly, *stc2* was also down-regulated in rare NCCs, a previously unknown site of expression. The specific pattern of *stc2* expression in two patches within the otic vesicle, possibly corresponding to the positions at which the sensory hairs form, makes this a potentially useful marker gene.

Stanniocalcin (*stc1*) regulates calcium uptake in the gills and small intestine and is the major antihypercalcemic hormone in bony fish (Ishibashi and Imai, 2002). A second stanniocalcin gene (*stc2*) has now been cloned in fish (Luo et al., 2005). A functional role for *stc2* has not been identified in fish but this study has identified *stc2* expression in the otic vesicle and a restricted number of NCCs for the first time. Thus the role of *stc2* in these tissues is of interest and could be examined through a morpholino approach. As *stc2* appears to be specifically expressed in two patches that correspond to the anterior and posterior maculae it is possible that this gene contributes to the development of the sensory hair cells at these sites. Indeed, these hair cell patches are expanded in *sox10* mutant otic vesicles (Dutton et al., 2009) and this corresponds with the expansion of *stc2* expression in mutant embryos. Other marker genes of sensory maculae patterning are also expanded in *sox10* mutant embryos (Dutton et al., 2009). Thus loss of function studies of this gene might be expected to unearth an anterior and posterior sensory maculae phenotype. It is intriguing that a gene can be both up-regulated and down-regulated in *sox10* mutant embryos, although in different tissues. Indeed it is perhaps surprising that such a gene was detected by microarray analysis. This behaviour is unique amongst all the targets of *sox10* identified during this screen.

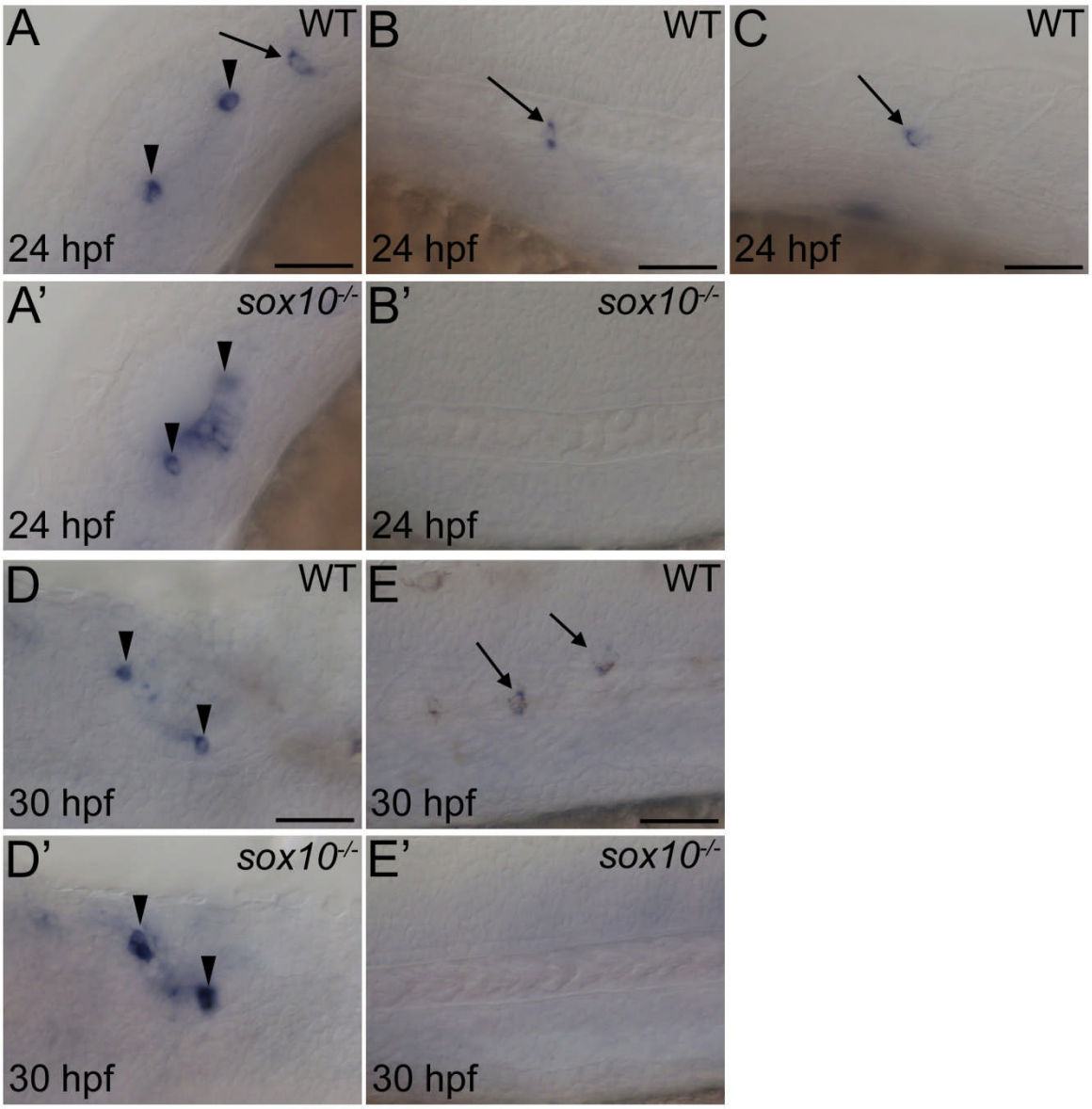


Figure 43: Whole mount *in situ* hybridization pattern of *stc2* expression.
All images are of zebrafish embryos at 24 hpf (A-C) or 30 hpf (D-E') treated. WT embryos are labelled with a letter alone, corresponding images of *sox10* mutant embryos are labelled with a letter and an apostrophe. Images A and A' are lateral views oriented anterior down and dorsal right, images D and D' are dorsal views of embryos oriented anterior left, all remaining images are lateral views orientated anterior left and dorsal top. Images A, A', D and D' are of the otic vesicle, all other images are views of the trunk. Expression of *stc2* in the otic vesicle is marked (arrowheads) and in cells outside of the otic vesicle (arrows), note expression on both the medial (B) and lateral (C) pathways. All scale bars = 50 µm.

4.2.6.2 sfrp5

Affymetrix Probe Identifier		Dr.21012.1.A1_at
Affymetrix Gene Symbol		*sfrp5*
ZFIN Gene Symbol		*sfrp5*
Full Name		secreted frizzled-related protein 5
Orthology		*SFRP5*
Biological Process GO Term		eye development multicellular organismal development Wnt receptor signalling pathway forebrain development
Molecular Function GO Term		Protein binding
7.2sox10:GFP array results	*t*-statistic	6.4778895378
	p-value	0.0029267701
	CLEAR result	diff.expr
	Fold Change	1.38
4.9sox10:GFP array results	*t*-statistic	0.1465975344
	p-value	0.8968923688
	CLEAR result	non.significant
	Fold Change	1.03
7.2+4.9sox10:GFP array results	*t*-statistic	1.1840668917
	p-value	0.2703716457
	CLEAR result	non.significant
	Fold Change	1.19

Table 43: Summary of information for *sfrp5*.

The Affymetrix probe set Dr.21012.1.A1_at corresponds to the gene *sfrp5*. This gene was identified as differentially regulated in *sox10* mutant embryos by both statistical tests applied to the 7.2 data set (**Table 43**). *Sfrp5* was not identified as differentially regulated in the other two data sets by either statistical test. Thus it was necessary to validate the 7.2 microarray data by *in situ* hybridization (**Figure 44**). *Sfrp5* expression at 24 hpf was visible but very weak in the otic vesicle (data not shown), it was not possible to genotype embryos by *sfrp5* expression at this stage. At 30 hpf, in WT embryos, *sfrp5* expression was observed in the otic vesicle in a ventro-medial position (**Figure 44A** and **B**). This otic vesicle expression pattern is similar to that already published for *sfrp5* on the ZFIN website (Thisse et al., 2004). The same pattern of expression was observed in *sox10* mutant embryos but expression levels were visibly increased in mutant embryos in comparison to WT embryos (**Figure 44A'** and **B'**). In summary, *sfrp5* is expressed in the otic vesicle and is up-regulated in *sox10* mutant embryos.

The sfrp family of genes code for secreted molecules that sequester Wnt signalling molecules in the extracellular space and thus are capable of antagonising both canonical (β-catenin dependent) and non-canonical (β-catenin independent) Wnt signalling (Kawano and Kypta, 2003, Li et al., 2008). Sfrp5 is known to play a role in modulating Wnt11 signalling during Xenopus gut development (Li et al., 2008). This study identified that local Sfrp5 inhibition of canonical Wnt11 signalling in the endoderm was required to promote

foregut identity. In addition, Sfrp5 inhibition of non-canonical Wnt11 signalling was essential for correct foregut morphogenesis, Sfrp5 depleted embryos displayed a lack of gut cell apical-basal polarity and cells were only loosely adherent (Li et al., 2008). A role for *sfrp5* during otic vesicle development has not yet been elucidated but it is plausible that sfrp5 modulates Wnt signalling in the otic epithelium to promote correct morphogenesis.

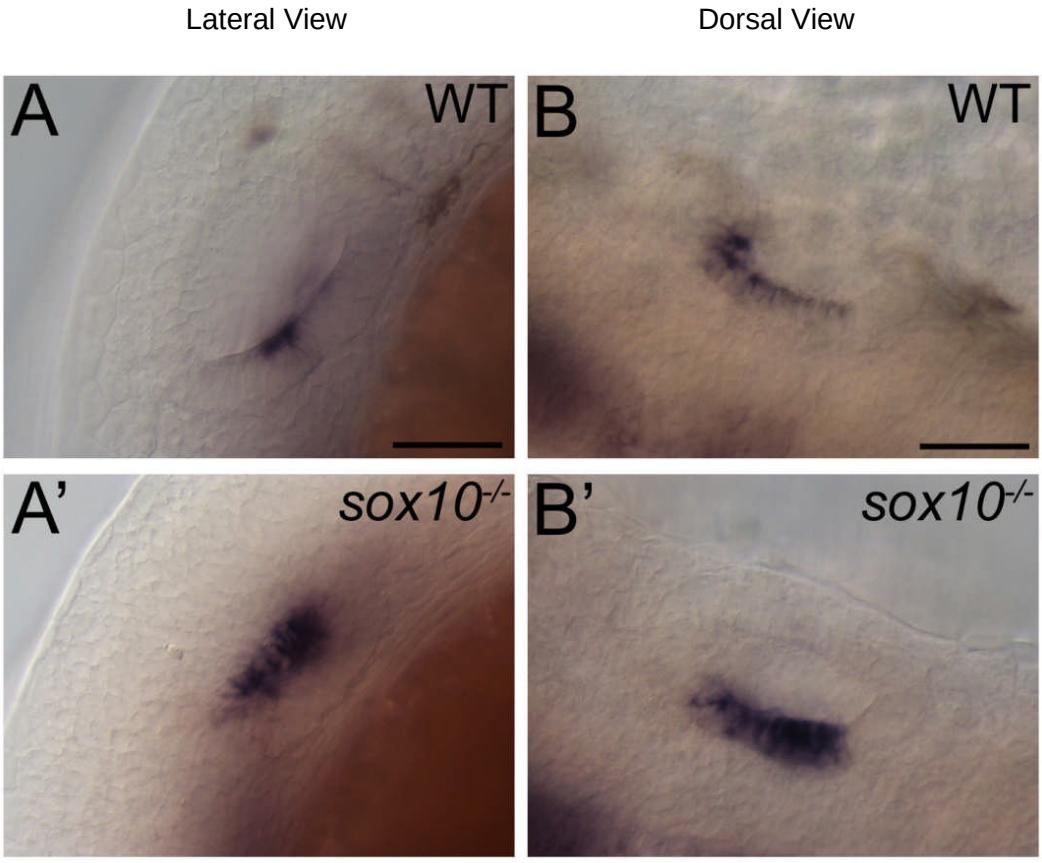


Figure 44: Whole mount *in situ* hybridization of *sfrp5* expression.
All images are of zebrafish embryos at 30 hpf showing the otic vesicle. WT embryos are labelled with a letter alone, corresponding images of *sox10* mutant embryos are labelled with a letter and an apostrophe. Images A and A' are lateral views of embryos oriented anterior down and dorsal left, images B and B' are dorsal views of embryos oriented anterior left. *Sfrp5* is expressed in a ventro-medial patch of the otic vesicle in WT and mutant embryos. All scale bars = 50 µm.

4.2.6.3 wnt11r displays possible NC expression

Affymetrix Probe Identifier		Dr.8297.1.S1_at
Affymetrix Gene Symbol		*wnt11r*
ZFIN Gene Symbol		*wnt11r*
Full Name		wingless-type MMTV integration site family, member 11, related
Orthology		*WNT11*
Biological Process GO Term		Wnt receptor signalling pathway, calcium modulating pathway multicellular organismal development heart development Wnt receptor signalling pathway cell migration involved in gastrulation
Molecular Function GO Term		signal transducer activity
7.2sox10:GFP array results	*t*-statistic	7.6103687286
	p-value	0.0016000068
	CLEAR result	diff.expr
	Fold Change	1.59
4.9sox10:GFP array results	*t*-statistic	2.7982866764
	p-value	0.107503697
	CLEAR result	diff.expr
	Fold Change	1.70
7.2+4.9sox10:GFP array results	*t*-statistic	3.7948172092
	p-value	0.0052754763
	CLEAR result	1.63
	Fold Change	diff.expr

Table 44: Summary of information for *wnt11r*.

The Affymetrix probe set Dr.8297.1.S1_at corresponds to the gene *wnt11r*. This probe set was identified as up-regulated in all three data sets by both statistical tests except the *t*-Test on the 4.9 data set (**Table 44**). To validate if this gene was up-regulated in *sox10*$^{-/-}$ embryos, *in situ* hybridization was performed (**Figure 45**). Expression of *wnt11r* was observed in the dorsal neural tube (arrowheads in **Figure 45**) consistent with NCCs prior to delamination. *Wnt11r* expression was also observed in a position consistent with pre-migratory NCCs post-delamination (arrows in **Figure 45**). *Wnt11r* is expressed in NCCs that contribute to the dorsal fin mesenchyme of *Xenopus* embryos (Garriock and Krieg, 2007), the *Xenopus* expression pattern bears a striking resemblance to the pattern observed here (Garriock and Krieg, 2007). However, no clear difference in the expression level of *wnt11r* was visible between WT and mutant embryos as determined by *in situ* hybridization. Given that all down-regulated NC expressed genes identified during this screen have shown a clear expression difference by *in situ* hybridization this may indicate that this gene is not NC expressed. If the gene is not expressed in NCCs no difference in gene expression would be expected, the result could be a false positive. In line with this, zebrafish *wnt11r* expression has been visualised in the same locations shown in **Figure 45** but the sites of expression were identified as the neural tube and dorsomedial somite edge (Groves et al., 2005). *Wnt11r* is expressed in the dorsomedial region of somites in

Xenopus embryos as well (Garriock and Krieg, 2007). Consistent with this, the *wnt11r*^{fh224} mutant has no obvious pigment defect although other NCC types may not have been assessed (C. Moens personal communication). It is possible that *wnt11r* is only expressed in NCCs in the dorsal neural tube prior to delamination. To resolve if *wnt11r* is expressed in zebrafish NCCs a double *in situ* could be performed with a *sox10* probe to check for expression overlap. Alternatively embryos from one of the *sox10*:GFP lines could be antibody stained to visualise GFP and combined with an *in situ* for *wnt11r* expression, again to assess if expression overlaps with identified NCCs. Alongside these experiments it would be necessary to determine if *wnt11r* expression is up-regulated in *sox10* mutant embryos using a more sensitive technique such as quantitative RT-PCR. This could be performed on FACS enriched NCCs from both WT and mutant embryos for comparison. To conclude, *wnt11r* is expressed in positions consistent with NCCs but differential regulation of this gene in *sox10* mutant embryos could not be validated by *in situ* hybridization. Further work to validate if this gene is up-regulated in the NC should be a priority as *wnt11r* is the only candidate in this category.

The expression pattern of *wnt11r* obtained here indicates that this gene is NC expressed. Interest in *wnt11r*, despite not being validated during this screen, has been sustained by research undertaken in *Xenopus* embryos. It was shown that *wnt11r* was expressed in NCCs that contributed to the dorsal fin mesenchyme (Garriock and Krieg, 2007). The authors identified that *wnt11r* signalling, via a Ca^{2+} dependent non-canonical Wnt signalling pathway, was required cell autonomously for the delamination of NCCs. Although fin mesenchyme NC contribution is restricted to skeletogenic NC (therefore not Sox10 dependent NCCs), the detection of this gene as up-regulated in the microarray data presented in this study hinted at a role for *wnt11r* in non-skeletogenic NC. A study in *Xenopus* has demonstrated that *wnt11r* is essential for cranial NCC migration (Matthews et al., 2008). But the authors identified that *wnt11r* was not expressed in the NC but in ectodermal cells adjacent to the NC. *Wnt11r* functioned non-cell autonomously and was essential for NCC migration. NCCs in morphant embryos could not delaminate from the neural tube. Thus this study and the study from Garriock and Krieg, 2007 identified similar functions for *wnt11r* in NCC migration but via different mechanisms. The published data is therefore contradictory regarding the sites of *wnt11r* expression. Comparable studies have not been performed in zebrafish but it would be very interesting to examine derivates of the NC in *wnt11r* mutants, focusing primarily on a defect in cell migration. Cell transplantation studies, similar to those performed in *Xenopus* embryos (Garriock and Krieg, 2007, Matthews et al., 2008) would be required to address if *wnt11r* functions cell autonomously or not. In summary, *wnt11r* was an intriguing result in the microarray data and the only possible up-regulated gene in NCCs, therefore *wnt11r* should be considered for further study to shed light on a possible role in NC migration.

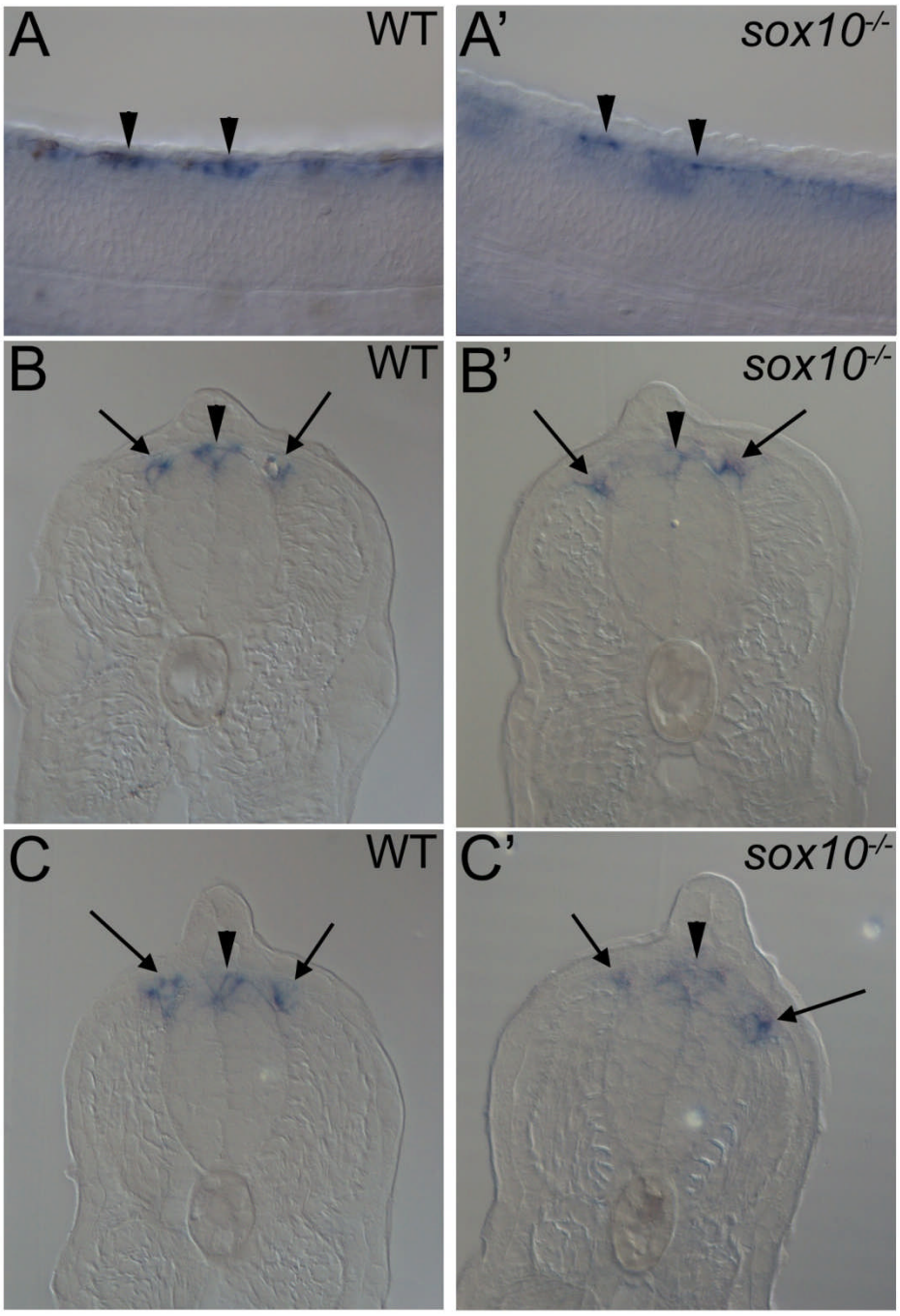


Figure 45: Gene Expression pattern for *wnt11r*.
All images are of 30 hpf embryos treated by *in situ* hybridization. Images A and A' are lateral views of trunk oriented anterior left and dorsal top, B-C' are images of transverse sections through trunk regions oriented dorsal top. All images labelled with a letter alone are of WT embryos, images labelled with a letter and apostrophe are corresponding images of *sox10* mutant embryos. *Wnt11r* expression was observed in putative NCCs in pre-migratory dorsal neural tube positions (arrowheads) and delaminated NCC pre-migratory positions (arrows).

4.2.7 Summary of the probe sets that were not validated.

During the *in situ* hybridization screen, a considerable number of genes expressed in the NC or otic vesicle were validated. However, the majority of genes screened were not expressed in these tissues. Many genes screened did not display obvious patterns of gene expression, displayed an expression pattern that was not spatially restricted or showed high levels of background that could not be resolved (**Figure 46**). Such genes were described as displaying non-specific expression. In both the down and up-regulated screens a large number of these genes were detected. In many cases these genes had also been described in a similar fashion on the ZFIN website. One example was the down-regulated gene *mkks*, no obvious expression pattern was determined for this gene and the same result had been published on ZFIN (Thisse et al., 2004). A minority of probes had a published expression on ZFIN that was not detected during this screen. For example, the up-regulated gene *pvalb7* is expressed in several sites including the hypaxial part of the anterior somites (Thisse et al., 2004) but this was not detected during the screen. This suggests that on occasion probes generated by the pipeline implemented during this screen were of a poor quality. Ten of the genes with no obvious expression pattern were only annotated with an Affymetrix identifier. Therefore only a limited amount of sequence data (typically between 300 to 400 base pairs) was available for probe design. It is perhaps not surprising that these probes were typically unable to specifically detect the distribution of mRNA.

During the screen it was observed that a large number of genes identified by the microarray were expressed in the central nervous system (CNS) or in the somites (**Figure 46**). These are not sites of *sox10* expression and these genes were not differentially regulated in *sox10* mutant embryos. However, ectopic GFP expression is present in the CNS and somites for both of the *sox10*:GFP transgenic zebrafish lines used during this project. It is tempting therefore, to speculate that both somite and CNS cells were sorted by FACS and therefore genes expressed specifically at these locations would be identified as expressed on the microarrays. By chance alone, the expression levels could have varied between WT and mutant samples thus resulting in these genes being identified as differentially regulated and subsequently screened by *in situ* hybridization. However, some genes were identified as differentially regulated that when examined by *in situ* hybridization displayed expression in other sites, for example the hatching gland, mucous secreting cells or the pectoral fin buds. These tissues do not display GFP expression in *sox10*:GFP transgenic embryos. Thus these genes have been identified as differentially regulated by chance.

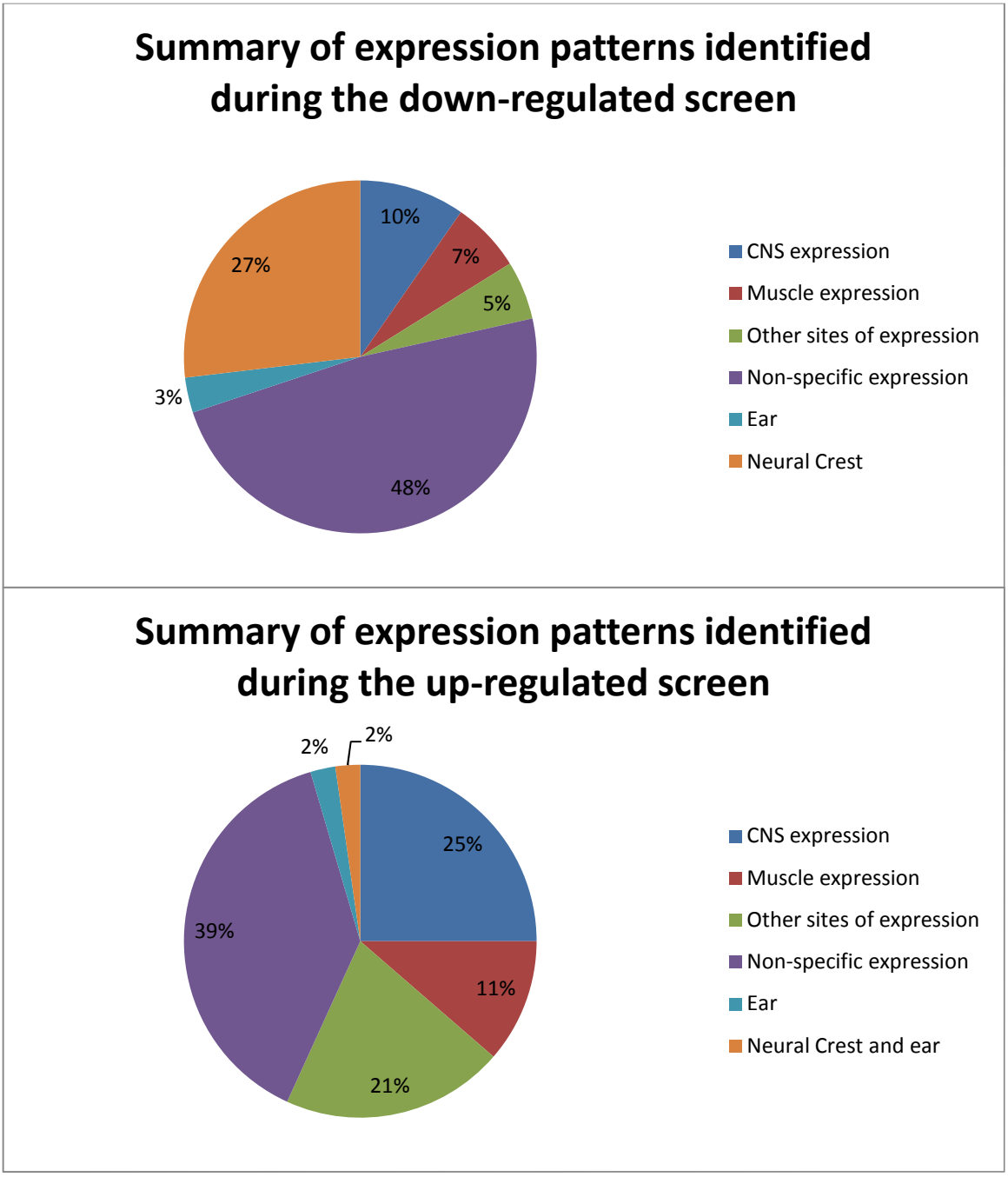


Figure 46: Pie charts summarising the sites of expression detected during the *in situ* hybridization screen.

4.3 Discussion

4.3.1 Further Work

Most of the genes identified as differentially regulated in *sox10* mutant embryos have clear expression patterns. This has enabled a strong prediction to be made about the cell type the gene was expressed in. This is not true in all cases and some genes show expression in cells not fully consistent with expression in a single cell type. For example, *cx33.8* shows a strong xanthophore lineage expression pattern but rare cells on the

medial pathway expressing *cx33.8* deviate from this. To clarify the cell types that each gene is expressed in, *in situ* hybridization could be performed comparing WT embryos with those of well known mutants. These mutant embryos lack or show defects in specific cell types, for example melanocytes (*mitfa*^{nacre}), xanthophores (*pfeffer* or *csf1r*^{panther}), iridophores (*ltk*^{shady}), glia (*foxd3*^{m188}) and sensory neurons (*neurog1*^{hi1059} or *sox10*^{baz1}). Thus if *cx33.8* medial pathway expression was absent in *ltk* mutant this would point towards expression in the iridophore lineage alongside the clear xanthophore expression. The validated miscellaneous genes, in particular *coro1c* and Dr.3972.1.S1_at should be prioritised for this type of evaluation. It might also be interesting to examine expression of *wnt11r* in different NC mutant embryos. In addition, further analyses of temporal and spatial expression patterns would be informative. For example, a time course of gene expression patterns could be produced for each gene. This could be combined with double *in situ* hybridizations with key marker genes for comparison. This would clarify when and where a gene is expressed and might help to indicate a possible function. These analyses would comprehensively characterise the gene expression patterns for the genes identified during this project and would ensure their use as valuable marker genes.

The majority of genes validated during the *in situ* hybridization screen have little or no function ascribed to them, particularly in regard to a role in NC or otic vesicle biology. Future work using morpholino mediated gene knock down or inhibitor studies should be focused on and would help to elucidate the function of these genes. Such a study could initially target genes with a very strong NC expression pattern, for example *cx33.8, slc2a15b, zgc:110343* or *zgc:100919*. Alternatively genes known to play a role in key NCC processes taking place at around 24 hpf such as specification, differentiation, migration or proliferation but have only been associated with the NC through this study could be prioritised. For example, both *degs1* and *zgc:100919* have been implicated in regulation of cell motility and cell proliferation. Alternatively genes with known functions but no clear functional relationship to NC biology could be investigated, for example, *aldh2b* (aldehyde metabolism), *zgc:110343* (reducing oxidative stress) or *cx33.8* (cell to cell communication). Such functional studies in zebrafish would be very informative regarding NC biology.

A large number of known NC expressed genes including *gch*, *aox3* (*xdh*) and members of the IMP biosynthesis pathway such as *gart* have been identified as differentially regulated by the microarray data but have not been validated by *in situ* hybridization. Thus there is clearly scope to continue the blind *in situ* hybridization screen which will lead to further known and novel genes being validated. An alternative approach would be to focus on particular classes or families of genes or genes with known functions of interest that have been identified as differentially regulated, for example transcription

factors. A focused *in situ* hybridization screen approach would also be likely to identify new targets of interest but would exclude genes that currently have little or no annotation to suggest a link with NC biology (See Chapter 5).

4.3.2 Regarding the Melanocyte Lineage Expressed Genes

4.3.2.1 Summary of the Melanocyte Lineage Expressed Genes Identified

The *in situ* hybridization screen has validated 8 genes that are expressed in the melanocyte lineage. Three of these genes, *dct*, *mitfa* and *pah*, are known marker genes of this lineage and thus their identification in the microarray analysis improved confidence that the screen has identified *in vivo* endogenous gene expression changes. The two V-ATPase subunit genes validated have links to melanocyte biology through the phenotype of insertional mutants (Amsterdam et al., 2004) but no gene expression pattern has been published. Interestingly more V-ATPase genes have been identified as differentially regulated by the microarray analysis and are therefore attractive candidates for validation. The three remaining genes, *zgc:100919, degs1* and *zgc:110239* have published neural crest gene expression patterns although they were not specifically annotated as marking the melanocyte lineage (Thisse et al., 2004). However there is no published data regarding the role these genes play in melanocyte development thus making them interesting candidates for further analysis, for example morpholino mediated gene expression knock down. Additionally *zgc:100919* has a very strong melanoblast expression pattern making it a valuable marker for this cell type.

4.3.2.2 V-ATPases and Neural Crest Development

Nine separate V-ATPase subunit genes were identified as differentially regulated in one or more of the microarray data sets (**Table 22**). Two of these genes, *atp6v1a* and *atp6v1e1*, were validated during the *in situ* hybridization screen. Both of the validated genes displayed a melanocyte lineage expression pattern with expression strongly reduced in pre-migratory NCCs and expression absent in migrating cells of mutant embryos. The validation of two V-ATPase genes makes the remaining candidates attractive genes for examination by *in situ* hybridization.

V-ATPase genes have been linked to NC biology by other studies. As a result of the insertional mutagenesis screen performed to identify genes important in zebrafish development, seven V-ATPase mutants were identified (**Table 45**). All of these mutants have a pigment phenotype showing a strong reduction in melanised cells, some also display an ear phenotype (Amsterdam et al., 2004). Fortuitously a mutant for *atp6v1e1*, one of the validated genes identified here, exists; this mutant has both a melanocyte and

an otic vesicle phenotype. The melanocyte phenotype corresponds well with the expression pattern identified during the *in situ* hybridization screen. In contrast, no otic vesicle expression was detected despite the small otolith phenotype of this mutant. The gene expression pattern detected for *atp6v1e1* was quite weak and had to be stained for an extended period of time. This led to the development of some background, particularly in the head, thus otic vesicle expression may have been obscured. Alternatively otic vesicle expression could be below detectable levels. This would correspond with the lack of otic vesicle expression detected not only for *atp6v1a*, but also for *atp6v1e1*. With these genes expressing subunits of the same proton transporter identical expression patterns would be expected. The majority of the V-ATPase mutants display a reduced head phenotype (**Table 45**) therefore it is feasible the otic vesicle phenotype of *atp6v1e1* is a secondary consequence of this consistent defect. Indeed the *atp6v1e1* mutant displays some very broad defects such as thin body, small head and underdeveloped gut that do not correspond with the expression pattern observed during this study. The mutant of *atp6v0c* has also been phenocopied by morpholino mediated gene knockdown (Pickart et al., 2004). The similarity of V-ATPase insertional mutants with the class VI.B pigmentation mutants *landing*, *slow tan* and *touch-down* has led to speculation that these mutants may have lesions in components of the V-ATPase complex (Kelsh et al., 1996, Pickart et al., 2004). Five V-ATPase subunit proteins were identified as components of melanosomes in a study designed to elucidate the melanosome proteome (Basrur et al., 2003). Thus components of the V-ATPase complex have a strong connection to melanocyte pigmentation which has been reinforced by the prevalence of the V-ATPase genes in this microarray screen.

The role of the V-ATPase complex in melanosome biology has yet to be fully understood. The V-ATPase complex is known to transport H^+ into intracellular compartments in an ATP dependent manner. This process acidifies the intracellular compartment and thus helps to regulate pH dependent processes (Nishi and Forgac, 2002). The pH of lightly melanised melanosomes (pH 4.6) and heavily melanised melanosomes was assayed (pH 3.0) and found to be acidic (Bhatnagar et al., 1993). This observation led to the assumption that a low pH was required for melanin biosynthesis. In this regard V-ATPases have an obvious role in ensuring the environmental conditions of the melanosome are suited to melanin biosynthesis. In contrast to this, the optimal pH for mammalian tyrosinase (the key enzyme in melanin biosynthesis) was suggested to be at an approximately neutral pH (Saeki and Oikawa, 1978, Ancans et al., 2001). Despite this, a mutation in a component of the V-ATPase complex leads to a reduction in pigmentation thus indicating a positive role in pigmentation during normal melanocyte development (Amsterdam et al., 2004). The application of V-ATPase inhibitors to melanoma cell lines increased tyrosinase activity and melanin production (Ancans and Thody, 2000). Thus the

role of the V-ATPase complex in melanocyte development is unclear, apparently having both positive and negative influences. This contradiction is reconciled if the insertion mutants actually up-regulate V-ATPase activity however this is unlikely as deletion of any gene that encodes a V-ATPase subunit in yeast leads to the entire complex failing to assemble (Nishi and Forgac, 2002). It is possible that during melanoblast development melanosomes require a low pH but at later stages during melanocyte differentiation, melanin biosynthesis is optimal at a neutral pH. Thus the difference in effects seen when knocking down V-ATPase function may be resolved by careful analysis of temporal function. Consistent with this, expression of *atp6v1a* was qualitatively stronger at 24 hpf than at 30 hpf. In line with this hypothesis it has been postulated that a switch in pH from acidic to neutral is required during melanosome maturation (Schallreuter et al., 2008).

Gene Name	Mutant Identifier	Phenotype
V-ATPase Ac45 or *atp6ap1*	hi112	day 2: no/less body pigment; day 5: less body pigment, small head and eyes, underdeveloped liver/gut
v-ATPase membrane sector associated protein M8-9 or *atp6ap2*	hi3681	day 2: no pigment, slightly smaller head and eyes; day 5: still no pigment, a little cns necrosis, slightly smaller head, eyes, and ears, a little pericardial oedema, underdeveloped liver/gut
v-ATPase subunit e or *atp6v1e1*	hi577a	day 2: little/no body pigment, day 3 spotty body pigment, small head, small otoliths, day 5: little/no body pigment, small head, eyes, and ears, underdeveloped gut, thin body
v-ATPase SDF/54kD subunit or *atp6v1h*	hi923	day 2: no/less body pigment; day 5: less body pigment, small head and eyes, underdeveloped liver/gut, sometimes some pericardial oedema
v-ATPase 16 kDa proteolipid subunit or *atp6v0c*	hi1207	day 2: no pigment and a little cns necrosis; d5: less pigment, also small head and eyes, underdeveloped liver/gut, sometimes bent and/or dying
v-atpase subunit F or *atp6v1f*	hi1988	day 2: no body pigment, less eye pigment; d5: less pigment, small and blocky head with some necrosis, small eye, underdeveloped liver/gut, a little pericardial oedema
v-atpase AC39 subunit or *atp6v0d1*	hi2188B	day 2: no/less body pigment, head a little shorter/blockier; day 5: less body pigment, body a little thinner, gut slightly underdeveloped

Table 45: V-ATPase subunit and V-ATPase associated genes with identified insertion mutants.
All information is taken from data published regarding the insertional mutagenesis screen (Amsterdam et al., 2004). Phenotypes are described in comparison to WT embryos.

4.3.3 Regarding the Xanthophore Lineage Expressed Genes

4.3.3.1 Summary of the Xanthophore lineage genes identified

From the 100 most down-regulated genes, of which 89 were examined by *in situ* hybridization, 12 probe sets represented genes expressed in the xanthophore lineage. These 12 probe sets corresponded to 9 separate genes (**Table 17**). In comparison to melanocytes, xanthophores have been poorly studied and have fewer well known markers. The well known xanthophore markers *gch* and *xdh* have both been identified as down-regulated in the microarray screen but do not feature in the 100 most down-regulated genes. The only well characterised xanthophore marker identified and validated in this screen was *pax7* (Minchin and Hughes, 2008). This paper identified *pax3* (Affymetrix probe identifier Dr.579.1.S1_at) as a key transcription factor in xanthophore development. This probe set was not identified as differentially expressed in any of the microarray data sets suggesting *pax3* expression is independent of Sox10. Six of the genes identified and validated (*rbp4l*, *atic*, *adsl, paics, zgc:110343* and *aldh2b*) have already been characterised as NC expressed but have no known functional role. Three of these genes are involved in the *de novo* IMP biosynthesis pathway and are likely to be important in generating GTP for pteridine biosynthesis. The two remaining genes have not been shown to be expressed in the NC or associated with a role in NC biology and are thus novel targets. *Slc2a15b* functions as a transporter and *cx33.8* may play a role in gap junction mediated cell to cell communication. Several of these genes, most noticeably *slc2a15b, cx33.8, paics* and *zgc:110343*, are highly expressed in the xanthophore lineage and therefore are likely to prove to be valuable markers in future studies of this cell type.

4.3.3.2 Connexins, sox10 and the neural crest.

The *cx33.8* expression pattern and lack of published material on the role of *cx33.8* in neural crest biology makes this an attractive target for further work. In addition there is a considerable amount of literature regarding the role of connexins; some research has linked them with both neural crest and *sox10* biology. Connexins assemble into hexamers called connexons (hemichannels), these are inserted into the plasma membrane. When a connexon in the membrane of one cell comes into contact and interacts with a connexon in an adjacent cell a gap junction is formed. This intercellular channel connects the cytoplasm of two cells and allows the exchange of ions, secondary messengers and small metabolites thus allowing electrical and biochemical communication between cells (Meşe et al., 2007). Evidence is also starting to accrue to suggest that unpaired connexons are functional and also that connexins have a function independent of their channel activity (Goodenough and Paul, 2003, Stout et al., 2004). Connexins and gap junctions function in

processes important in NC development including cell differentiation and migration (Kojima et al., 2008, Xu et al., 2006).

Connexins are known to play roles in NC biology. A zebrafish adult pigment patterning mutant, *leopard*, is caused by lesions in *cx41.8* (Johnson et al., 1995, Watanabe et al., 2006). The *leopard* mutant has all pigment cell types but during adult pigment pattern formation, pigment cells become organised into spots rather than stripes (Johnson et al., 1995). The precise function of *cx41.8* in pigment cell patterning has yet to be elucidated but the *leo*^{tq270} mutant displays a strong phenotype and disrupted gap junction activity (Watanabe et al., 2006). In line with this, transplantation experiments identified that homotypic and heterotypic interactions between melanocytes and xanthophores are required to establish the adult zebrafish stripe pattern (Maderspacher and Nüsslein-Volhard, 2003). Thus cell to cell communication mediated by gap junctions is important during NC patterning.

NCCs undergo extensive migration throughout the embryo to various locations, during this process connexins are up-regulated (Kuriyama and Mayor, 2008). The expression of *Cx43* has been documented in mouse trunk NCCs by both *in situ* hybridization and through analysis of a transgenic mouse in which a LacZ reporter was under control of the *Cx43* promoter (Lo et al., 1997). Experiments *in vitro* demonstrated that NCCs over expressing *Cx43* migrated faster than cells from a *Cx43* knock out mouse. This effect was also observed *in vivo* using the *Cx43-LacZ* transgenic mouse; more LacZ positive cells from embryos over expressing *Cx43* migrated to their final destination in the heart than LacZ positive cells in *Cx43* knock out embryos (Huang et al., 1998). Thus *Cx43* appears to play a role in the migration of cardiac NCCs. Further studies have indicated that this role may be independent of gap junction communication but dependent on interactions with the actin cytoskeleton (Xu et al., 2006, Xu et al., 2001). A further possible role of connexins in the NC was indicated by the inhibition of gap junction communication using chemical inhibitors (uncoupling agents). This reduced the survival of rat trunk NCCs *in vitro* with weaker inhibitors quantitatively inducing a weaker effect (Bannerman et al., 2000). Thus connexins have been associated with several aspects of NC biology including pigment cell patterning, migration and cell survival. However no role for connexins and gap junctions has been identified in early embryonic pigment cells. Thus, *cx33.8* identified in this study is so far unique by being clearly expressed in a pigment cell type and may function in one of the roles associated with connexins. The strong expression of *cx33.8* in xanthophores suggests that this gene might regulate cell to cell communication between xanthophores. The alternative is that *cx33.8* may have a function as a hemichannel in xanthophores or have a function independent of gap junction activity.

In addition to the roles that connexins play in the NC, this family of genes is also of interest because they are known to be regulated by Sox10. Co-transfection of the *Cx32* promoter up-stream of a reporter gene with a human SOX10 expression construct identified *in vitro* regulation by *Sox10* of a connexin gene. The study also identified SOX10 binding sites in the *Cx32* promoter suggesting that this regulation may be direct (Bondurand et al., 2001). In a similar transient co-transfection experiment, SOX10 regulation of *Cx32* and *Cx47* was identified (Schlierf et al., 2006). SOX10 binding sites were identified in the promoter sequence of *Cx47*, again showing that this regulation could be direct. While there is no evidence that *cx33.8* is a direct target of Sox10 this could be investigated by assessing the promoter for Sox10 binding sites and through co-transfection experiments.

Study of connexin genes, particularly with regard to *cx33.8*, is likely to prove amenable in the zebrafish model organism. Single cell labelling can be performed in zebrafish embryos and is often used to trace the fates adopted by the progeny of a cell (Dutton et al., 2001a). To study cell to cell coupling mediated by gap junctions, a single NCC could be injected with neurobiotin or another low molecular weight marker. Such a dye can pass between coupled cells to demonstrate functional gap junction communication (Bannerman et al., 2000). It would be exciting to observe if this communication occurs between NCCs. Uncoupling agents, such as 18α-glycyrrhetinic acid, which inhibit the gap junction function of all connexins could also be used on zebrafish embryos (Bannerman et al., 2000). As these agents are not specific inhibitors of a particular connexin, they would need to be used with caution perhaps limiting application to a specific period of time. It would be relevant to determine the full complement of connexin genes expressed in NCCs during zebrafish development. Inhibition of a specific connexin could also be studied in zebrafish using a morpholino approach. A morpholino approach was adopted during this project to examine the function of *cx33.8* in the zebrafish NC. This was a very preliminary experiment and as such has not been presented fully within this thesis. Briefly, c*x33.8* morphant embryos (9.2 ng injected) appeared to have fewer *gch* positive cells than control (phenol red injected) embryos. Counts of *gch* positive cells migrating in morphologically normal 30 hpf WT (phenol red injected) and *cx33.8* morphant embryos (*n*=15 for control group, *n*=18 for morphant group) were performed. Cells were counted migrating over the 8 somites anterior of the anus and only cells separate from the dorsal stripe were counted. This showed a statistically significant difference between the two groups as determined by a two tailed *t*-Test. Averages of the cells counted per embryos from the two groups are presented with standard deviation; control average = 22.9 ± 4.3, morphant average = 14.4 ± 5.3. It was not determined if the reduction in *gch* positive cells was the result of xanthophore migration failing, being delayed or as a result of a reduction in xanthophore numbers.

More comprehensive counts and more carefully controlled experiments, including using a mismatch and a second non-overlapping morpholino, are required to validate and further examine this observation.

4.3.3.3 Regarding the de novo IMP biosynthesis pathway.

Three genes encoding enzymes of the *de novo* IMP biosynthesis pathway were present in the 100 most down-regulated genes in the *7.2sox10*:GFP microarray data set and were validated during the *in situ* hybridization screen. Several of the remaining genes in this pathway are also down-regulated according to the microarray data (**Table 32**) and have published NC expression patterns. These genes are therefore likely to be down-regulated in *sox10*$^{-/-}$ embryos. The identification of multiple components of this pathway by microarray analysis helps to both validate the screen and to suggest that this pathway functions during xanthophore development at around 24 hpf.

The connection between this pathway and xanthophore development has not been elucidated. The starting point in the synthesis of pteridines, the yellow pigment of xanthophores, is the purine nucleotide guanosine triphosphate (GTP) (Ziegler, 2003). The initial step in pteridine biosynthesis involves GTP being converted into H_2neopterin triphosphate by Gch. Thus a supply of purine nucleotides is likely to be important in pteridine synthesis and xanthophore development. To generate GTP from IMP requires the synthesis of xanthosine monophosphate (XMP) from IMP and this is catalysed by the enzyme IMP Dehydrogenase (IMPDH). The *impdh* (si:dkey-31f5.7 on ZFIN) gene is differentially expressed with high variability according to the CLEAR test on the 7.2 microarray data and is down-regulated (fold change = -1.51). According to *in situ* hybridization data submitted to ZFIN, *impdh* is expressed in the NC (Thisse et al., 2004). The reaction to synthesise guanine monophosphate (GMP) from XMP is catalysed by the enzyme guanine monophosphate synthetase (GMPS). According to *situ* hybridization data submitted to ZFIN this gene is also NC expressed (Thisse et al., 2004) however no Affymetrix probe set corresponds to this gene. Phosphate groups can then be added to GMP by kinase enzymes to synthesise GTP, the starting molecule of pteridine synthesis. Thus a supply of IMP is likely to be required as a substrate for pteridine synthesis and therefore for xanthophore pigmentation. IMP biosynthesis enzymes have been strongly represented in the microarray data and *in situ* hybridization validation presented here.

At later stages in development (2 dpf) many of the *de novo* IMP biosynthesis genes appeared to be expressed in scattered cells in the dorsal and ventral stripes. These positions are consistent with cells of the iridophore lineage. Thus it is possible that the *de novo* IMP biosynthesis pathway is required for both xanthophore and iridophore development. The iridescence of iridophores is a result of light reflecting plates within the

cell that are comprised of purine crystals (Ziegler et al., 2000, Rohrlich and Rubin, 1975). Thus the synthesis of purine nucleotides like IMP may be required to generate purine for crystallisation. It is interesting to note that *adsl* is expressed in rare medial pathway migrating cells, these could be iridoblasts. Consistent with this, only a few *ltk* expressing iridoblasts can be seen migrating on the medial pathway at 26 hpf (Lopes et al., 2008). As there are very few markers of the iridophores lineage available it would be worth examining *adsl* expression in *ltk* mutant embryos.

A key gene in the specification of iridophores is the receptor tyrosine kinase *ltk* (Lopes et al., 2008). *LTK* is most closely related to another receptor tyrosine kinase, *ALK* (Morris et al., 1997). In ALK positive anaplastic large cell lymphoma tumours chromosomal translocations juxtapose *ALK* to other genes. This generates ALK fusion proteins that have constitutive kinase activity, this perturbs cell signalling leading to a neoplastic phenotype. One of the known fusions that occurs to activate ALK is with the *de novo* IMP biosynthesis pathway enzyme *ATIC* (Trinei et al., 2000). However recent evidence has also identified ATIC as a direct phosphorylation target of ALK (Boccalatte et al., 2009). The authors discovered that the enzymatic activity of ATIC was enhanced by ALK dependent phosphorylation. Considering the close relationship between *ALK* and *LTK*, it is conceivable that *ltk* may play a role in *atic* regulation and *de novo* IMP biosynthesis in iridophores.

Cells that are rapidly proliferating require a supply of nucleotides, these can be supplied by either the *de novo* nucleotide biosynthesis pathways or by the nucleotide salvage pathways. Normal cells prefer the salvage pathway while cancer cells are known to rely heavily on the *de novo* purine biosynthesis pathway for ATP and GTP nucleotides (Li et al., 2007). As such a variety of inhibitors of enzymes in the *de novo* purine biosynthesis pathway have been developed, for a review see (Christopherson et al., 2002). Cells of the xanthophore lineage express several enzymes from this pathway (this study and Thisse et al., 2004) and characterised mutants of these enzymes display xanthophore defects (Amsterdam et al., 2004). Thus it seems likely that *de novo* purine biosynthesis is important in xanthophore development. With xanthophores being an easily visualised cell type in zebrafish, this may provide an amenable model to test the efficacy of IMP biosyntheis inhibitors in an *in vivo* context. To progress this idea, the identified mutants of *de novo* IMP biosynthesis genes need to be fully characterised to identify at what stage in xanthophore development the defect occurs. For example, do these mutants fail to specify xanthophores, do the cells fail to differentiate or do they have a normal morphology but simply do not pigment? The increased number of xanthophore markers available as a result of this microarray study will be a valuable tool in achieving this. As the *de novo* IMP biosynthesis pathway is likely to produce a substrate for pteridine synthesis, these mutants are likely to only have a defect in xanthophore pigmentation.

Once the defect has been examined, suitable *in situ* hybridization markers will be available to complement the visual assessment of xanthophore yellow pigmentation. Thus the efficacy of *de novo* IMP biosynthesis pathway inhibitors could be accurately examined. It is encouraging to note that apart from lacking yellow pigment, mutants of enzymes in the *de novo* IMP biosynthesis pathway are grossly normal and therefore toxic side effects of inhibitors would generate striking phenotypes. It should also be noted that using zebrafish as a model organism will enable medium to high throughput analysis of potential inhibitors.

4.3.4 Regarding the Differentially Regulated Otic Epithelium Expressed Genes

Three otic epithelium expressed genes have been identified and validated as down-regulated in *sox10* mutant embryos during this project. This represents a hit rate of 3/89 (3.4 %). Two of these genes (*cldnj* and *otomp*) have already been identified as otic epithelium expressed genes and analysis of mutant or morphant embryos has been conducted. As a result of functional studies it was determined that both *cldnj* and *otomp* knockdown results in otoliths forming but failing to grow properly. Thus both genes play a role in otolith development and loss of function experiments show that these genes are required for otoliths to grow at the correct rate. Despite the similar phenotype elicited by loss of function of these two genes, *otomp* and *cldnj* have different expression patterns. Both genes are expressed throughout the otic epithelium but *cldnj* expression is concentrated in the dorso-lateral region of the otic epithelium while *cldnj* expression is more intense in a ventro-posterior region. The other Affymetrix probe set validated (Dr.18158.1.A1_at) has not been assigned to any known transcript yet. The data presented here demonstrates that the portion of chromosome 21 that Dr.18158.1.A1_at corresponds to is transcribed. Again the pattern of expression of this gene is different to *otomp* and *cldnj*. Dr.18158.1.A1_at is chiefly expressed in a ventro-medial region with some expression extending to a ventro-lateral area. It would be very interesting to identify the gene at this chromosomal location and then to pursue functional studies. It would be fascinating to identify which characteristic of the complex *sox10* mutant otic vesicle is phenocopied by knock down of Dr.18158.1.A1_at. This result will contribute to unravelling the molecular basis of the zebrafish *sox10* mutant otic vesicle phenotype and may shed light on human deafness often associated with WS4 syndrome.

Two otic epithelium expressed genes (2/47 or 4.3 % hit rate) have also been validated as up-regulated in *sox10* mutant embryos, *sfrp5* and *stc2*. *Sfrp5* is expressed strongly in a ventro-medial region of *sox10* mutant embryos with weaker expression at the same sites in WT embryos. *Stc2* was expressed in two patches towards the ventro-medial region of the otic vesicle in WT embryos. Interestingly, expression of *stc2* was up

regulated in *sox10* mutant embryos at the endogenous WT sites but also displayed expression in ectopic sites. Careful examination of the phenotype of *sox10* mutant embryos determined that *sox10* mutant otic vesicles show morphological and patterning defects (Dutton et al., 2009). An examination of molecular markers of otic vesicle patterning such as *sox9a, sox9b, pax2a, fgf8, fsta, bmp4 and wnt4a* in WT and *sox10* mutant embryos was performed. It was discovered that these markers were up-regulated and often displayed ectopic expression, sometimes associated with a loss of expression at normal sites (Dutton et al., 2009). This could explain why the majority of these genes were not identified as differentially regulated in the microarray screen. It would be interesting to dissect how the altered regulation of all of these genes, including *stc2* and *sfrp5*, has contributed to the *sox10* mutant otic vesicle phenotype. This study has established that ear patterning genes show derepression and deregulation in *sox10* mutant embryos (Dutton et al., 2009). It is therefore not surprising that ear genes have been identified as up-regulated in the microarray screen. However it is not clear if Sox10 directly or indirectly regulates these targets.

4.3.5 Regarding the Down-Regulated Genes

During the *in situ* hybridization screen, 89 down-regulated genes were assessed, of these 25 were validated, a hit rate of 28 %. All genes that showed NC or otic vesicle expression were also down-regulated in *sox10* mutant embryos. The majority of these genes (17/25) appeared to be expressed in melanocytes and xanthophores and 3/25 were otic vesicle expressed. The neural crest marker gene *crestin* was also validated. Only one gene (*foxd3*) was expressed in the peripheral nervous system and three genes require further work to further characterise their NC expression pattern. No clear markers of the iridophore lineage were identified. It has recently been shown that iridoblasts specify at around 24 hpf (Lopes et al., 2008), the time point being examined by the microarray experiment. Thus iridophore specific genes down-regulated in *sox10* mutants were expected to be detected. Cells of the iridophore lineage only migrate on the medial pathway and their identification could be masked if a potential marker is also expressed in another medial pathway migrating cell type, melanocytes for example. Additionally, some xanthophore marker genes displayed rare cells on the medial pathway, for example *cx33.8* and *adsl,* these cells may represent iridoblasts. The xanthophore expressed IMP biosynthesis pathway genes, particularly *adsl*, are of extreme interest in this regard as they appear to be expressed in iridophores at 2 dpf. Analysis of the expression pattern of potential iridophore genes in the *ltk*^{shady} mutant would help to identify any iridophore lineage expression.

In regards to the scarcity of identified peripheral nervous expressed genes, neuronal and glial fated cells are migrating at around 24 hpf thus some degree of specification is likely to have occurred (Raible and Eisen, 1994). It is unclear to what degree these cells are specified and what molecular correlates of this are expressed at this stage. Therefore some neuronal and glial expressed genes might have been expected to be identified as down-regulated by microarray analysis. However these genes may not be dependent on Sox10 and thus would not be detected during this project. In addition, the mutant phenotype of neuronal cell types is less severe than for pigment cells in *sox10* mutants as some cells differentiate correctly. For example *neurog1* positive sensory neurons are only reduced in *sox10* mutant embryos, unlike melanocytes which are almost completely absent (Carney et al., 2006). At 24 hpf neurons in the cranial ganglia appear normal in *sox10* mutant embryos (Kelsh and Eisen, 2000) and may not display molecular differences between WT and mutant embryos at this stage. As a result, neuronal expressed genes may prove difficult to detect by microarray analysis. The enteric nervous system (Elworthy et al., 2005) and sensory neurons (Carney et al., 2006) are also specified later than 24 hpf therefore these cell types were not examined during this microarray screen. However, *foxd3*, which marks glial precursors, was detected as down-regulated by the microarray analysis, perhaps because the glial phenotype of *sox10* mutants is more severe than for sensory neurons (Carney et al., 2006). Therefore it would be expected that other glia expressed genes have been identified as down-regulated in the microarray screen and are awaiting validation by *in situ* hybridization.

Of the validated genes, 7 were completely novel to NC biology, a further 8 genes had no known functional role in the NC and one gene was not known to be expressed in the NC but had a pigment mutant phenotype. Thus the screen has been successful in validating 16 genes identified as down-regulated by microarray analysis that are at least to some extent novel. This has identified many interesting targets for further investigation. In summary, the microarray screen appears to have mainly targeted pigment cell specific genes and the *in situ* hybridization approach has proved successful in elucidating temporal and spatial gene expression patterns and validating down-regulation in *sox10* mutant embryos.

4.3.6 Regarding the Up-Regulated Genes

Two otic vesicle expressed genes were validated as up-regulated by *in situ* hybridization. This represents a hit rate of 2/47 or approximately 4 %, similar to the 3 % hit rate of validated down-regulated otic vesicle expressed genes. No other genes were observed as NC or otic vesicle expressed. *Sfrp5* displayed expression in a ventro-medial region of the otic vesicle that was stronger in *sox10* mutants when compared to WT embryos. *Stc2* was

expressed in two distinct patches in WT embryos, possibly marking two patches of sensory hair cells, in mutant embryos this expression was stronger and showed ectopic expression in the otic vesicle. Surprisingly *stc2* was also expressed in rare NCCs and was down-regulated in these cells in mutant embryos. Thus up-regulation of two novel ear specific genes was identified in *sox10* mutants and is likely a result of deregulation and derepression (see section **4.3.4**) (Dutton et al., 2009).

No clear example of a NC up-regulated gene was identified despite screening 47 of the top 50 most up-regulated genes. The iridophore marker gene *ltk* has been shown by *in situ* hybridization to be strongly up-regulated with increased numbers of *ltk* positive NCCs in *sox10* mutant embryos at around 24 hpf (Lopes et al., 2008). Thus up-regulation of gene expression does occur in *sox10* mutant NCCs. The Affymetrix zebrafish chip lacks a probe set that detects transcripts from the *ltk* gene therefore no positive control for up-regulated genes was available. This has made it difficult to interpret the lack of up-regulated NCCs identified by the microarray approach. In *sox10*^{m618} mutant embryos, the supernumerary *ltk* positive cells also expressed *sox10* (Lopes et al., 2008). This led to the hypothesis that these *ltk* positive cells, in line with the fate restriction model of NC development, represent NCCs trapped in a partially restricted progenitor phase. Cells are trapped because without functional *sox10* they cannot become specified to individual derivative cell fates (Dutton et al., 2001a). It was predicted that such trapped cells would show up-regulated expression of genes that mark an intermediate phase in NC development, for example a chromatoblast precursor cell. A lack of such examples in the array data does not exclude their existence; it does raise the question why these examples were not identified. This can be viewed in contrast to the successful identification of up-regulated otic vesicle genes. *Sox10* plays a key role in the specification of non-skeletogenic NC derivatives by regulating key transcription factors thereby activating a cascade of genes required for the development of a specific cell type (Dutton et al., 2001a, Carney et al., 2006, Elworthy et al., 2003, Elworthy et al., 2005). Thus without functional *sox10*, the expression of a large number of genes is defective as demonstrated during this project. In the otic vesicle, *sox10* plays a role in regulating genes involved in patterning and morphology (Dutton et al., 2009). A loss of *sox10* deregulates these processes leading to abnormal gene expression. As such up-regulation of otic vesicle genes in *sox10* mutants could have been predicted and has been demonstrated (Dutton et al., 2009). However, the example of *ltk* demonstrates clear and strong up-regulation of a NC gene in *sox10* mutants. It would be interesting to see if this up-regulation was maintained after FACS. It is possible that in *sox10* mutants, the trapped NC progenitor cells represent a highly heterogeneous population of cells with variable gene expression profiles. A *t*-Test tends to select against genes that show variable gene expression across samples. Thus it is possible that no clear pattern of up-regulated gene

expression would be detected by microarray analysis using a *t*-Test. The *in situ* hybridization pattern of *ltk* in *sox10* mutants shows that a large number of NCCs ectopically expressed *ltk* suggesting that this gene would have been identified as up-regulated with a large fold change. This argues that the trapped NCC population, even if highly heterogeneous, still strongly express certain genes. Therefore the CLEAR test, which assesses fold change, may have been a more powerful test than the *t*-Test to detect this. It is also possible that there are only a few genes that mark these trapped populations of cells and these genes are poorly represented on the Affymetrix array. An alternative hypothesis is that the disaggregation of embryos for FACS has disrupted the expression profile of these unspecified cells, perhaps by denying them specific secreted signalling molecules. Thus the unstable progenitor state expression profile would be lost while a stable specified cell fate expression profile in WT cells might not be so disrupted. The genes expressed in trapped NCCs in *sox10* mutants may represent examples of NCSC genes, markers of multipotency or receptors for instructive ligands and thus represent genes of extreme interest. It is therefore disappointing that such examples have yet to be characterised from the microarray data. However the value of this type of target ensures that further investigation of the data is worthwhile in an attempt to uncover them.

4.3.7 General Discussion

The *in situ* hybridization screen has validated 27/136 (20 %) of the assessed genes, in fact this hit rate increases to 30/136 (22 %) when duplicate probe sets are included. All the genes that showed clear expression in the cell types of interest were differentially expressed in *sox10* mutant embryos compared to WT embryos. The hit rate for down-regulated genes was 28 % and for up-regulated genes was 4 %. This represents a striking difference and it should also be noted that all the validated up-regulated genes were expressed in the otic vesicle. This suggests that *sox10* has a primary function in the NC to activate gene expression but may play a different role in the otic vesicle.

The *in situ* hybridization screen approach has the benefit of generating gene expression patterns often for genes with no published expression pattern. This has enabled genes to be characterised into groups based on gene expression pattern. A technique such as RT-PCR could not achieve this, indeed it could not easily distinguish between otic vesicle and NC expressed genes. Despite this advantage, an *in situ* hybridization screen which involved the cloning of most of the genes of interest has proved to be a labour intensive and costly process. These difficulties could be partly negated by collaboration with a group that has access to an extensive cDNA library for *in situ* hybridization probe generation. Examination of microarray data by RT-PCR tends to be able to validate more targets than an *in situ* hybridization approach, for example, Qian

et al., 2005 examined 13 genes validated by RT-PCR as differentially regulated in *cloche* mutant microarray data. Only 8 of these target genes showed expression in cell types of interest by *in situ* hybridization. The remaining 5 genes did not display a specific expression pattern. Thus *in situ* hybridization may be a less sensitive technique but equally provides a more informative and stringent test. Genes that did not show specific expression patterns were deemed not to be NC or otic vesicle expressed during this screen.

Analysis of the gene expression patterns of the validated genes determined that the majority of genes were expressed in melanocytes or xanthophores. This has greatly expanded the number of marker genes now available for these cell types. No clear markers of the iridophore lineage were discovered but several genes show potential in this regard and are worth investigating. Equally, only one well known glia marker and no clear neuronal markers were validated from the microarray data. With a paucity of markers available for the peripheral nervous system it would be worth investigating the potential PNS expressed genes *coro1c* and Dr.3972 to determine their expression pattern more fully.

A number of known NC expressed genes have been identified as differentially regulated in the microarray data but were not assessed during the *in situ* hybridization screen. It would be expected that these genes show differential expression in *sox10* mutant embryos in comparison with WT embryos. This demonstrates that further discoveries are waiting to be made from the microarray data. This could be achieved by further blind screening or by targeting gene types of interest, for example transcription factors. Alternatively, published expression patterns could first be used to interrogate a list of genes identified by microarray analysis to eliminate genes not expressed in the cell type of interest. These ideas will be explored in more detail in Chapter 5. In summary, this combined microarray analysis of *sox10* mutant embryos and *in situ* hybridization screen has been successful so far and retains the potential to facilitate further discoveries.

Chapter 5: Further Exploration of Microarray Data

5.1 Introduction

One challenge of a microarray project arises when selecting genes for validation. Typically a large number of candidates are designated as significantly differentially regulated. One option available to examine identified genes is to use published data, such as expression patterns on ZFIN. When well characterised expression patterns are available these could be used to identify if a gene is expressed in the tissue or cell type of interest or not. This approach will focus on genes that are already well characterised to reduce the number of genes requiring validation by eliminating false positives and to highlight likely true positives.

A major advantage of microarray data is that novel candidates can be identified that have not previously been associated with a biological process. This can identify genes for study that have little or no prior publication history. However these candidates are unknown entities and are therefore only likely to be identified during validation by methods such as a blind *in situ* hybridization screen. The screen performed in Chapter 4 was a blind screen that identified many interesting known and novel genes; unfortunately a large number of false positives were also assessed. This screen was based on the ranking of genes by *t*-statistic (with associated *p*-values used to determine significance). Alternatively, fold change has been a popular measure used to select genes of interest from microarray data. Fold change is the ratio of RNA quantities between two groups of samples in a microarray experiment (Cui and Churchill, 2003). Thus fold change is used as a measure of the change in gene expression levels between samples but is not associated with any level of statistical significance. Fold change was used to rank genes that were up-regulated during the induction of NCCs *in vitro* (Adams et al., 2008, Gammill and Bronner-Fraser, 2002). Thus this method may help to identify genes that are easily validated by a method such as *in situ* hybridization but without performing a statistical test there is no associated level of confidence in the designation of genes as differentially regulated or not differentially regulated (Cui and Churchill, 2003).

During the statistical analysis of a microarray experiment a test is carried out on each gene leading to thousands of tests being performed. This results in the accumulation of false positives, when a gene is declared as differentially regulated when in fact it is not (type 1 error). Methods are available to correct for multiple testing (also see Chapter 3). One approach to multiple testing controls the family-wise error rate (FWER) or the probability of accumulating false positive errors. FWER criteria are very stringent therefore lists of genes significant after controlling FWER are associated with high confidence that they do not contain any false positives. The reality is that while FWER significant genes

are likely to be significantly differentially regulated, a large number of genes will falsely be rejected and identified as not differentially regulated, false negative (type 2 error) (Cui and Churchill, 2003). This method of controlling multiple testing may not be appropriate when attempting to conduct a screen to identify a large number of differentially regulated genes. Controlling the false-discovery rate (FDR) represents an alternative strategy and defines the proportion of false positives among all of the genes identified as being differentially expressed (Cui and Churchill, 2003). At a defined FDR significance level a known proportion of genes should be false positives; this approach should help to ensure that true positives are not rejected. Thus at a significance level of $p<0.05$, 5 % of identified candidates are likely to be false positives. This represents a more liberal approach to controlling for multiple testing and FDR provides a sensible balance between the number of true positives and false positives in a microarray study (Storey and Tibshirani, 2003). Therefore a range of options are available to rank these significantly differentially regulated genes to facilitate selecting genes for validation. The investigator would aim to select the method which maximised true positives and minimised false positive results.

During this project two microarray data sets have been generated from the *7.2sox10*:GFP and *4.9sox10*:GFP transgenic lines. These two sets of samples are derived from different biological sources but both examine differential gene regulation in *sox10* mutant embryos compared to WT embryos. Thus any gene identified as differentially regulated in both data sets has been identified twice independently. It would be expected that the false positives (identified as differentially regulated by chance) would not be common between the two differentially regulated lists of genes. This analysis may therefore present an opportunity to eliminate false positives. The caveat is that not all of the marker genes known to be down-regulated in *sox10* mutants were identified as such in both data sets. In addition otic vesicle expressed genes may also be excluded as the *4.9sox10*:GFP transgenic line only weakly labels otic vesicle cells. As such this technique might also eliminate true positives that have been identified as differentially regulated during the *in situ* hybridization screen. These two samples generated from different transgenic lines have been treated as biological replicates by producing the combined data set to generate a set of 10 independent biological replicates (5 WT and 5 mutant samples). In doing so the number of microarray chips being evaluated has been increased. This increase in sample size may generate data that is more reliable than data from the two separate sets of arrays.

Selecting genes of interest based on function represents a third option for cherry picking genes of interest from microarray data. For example, genes annotated with an interesting gene ontology (GO) term or genes that are part of a gene family of interest such as zinc finger transcription factors could be selected. Affymetrix probe annotation on the NetAffx website provides both the gene name and functional annotation in the form of

GO terms thus facilitating this form of analysis. While this approach will focus on genes that have been studied previously, at least to some extent, it would be capable of identifying genes that have not been previously associated with biological process of interest. Selected genes would require subsequent validation with a number of candidates likely to be false positives.

Gene ontology or GO terms form a hierarchical structure of precisely defined, common and controlled vocabulary of terms for describing the normal roles of genes and gene products (Ashburner et al., 2000). GO terms have been split into three categories. A biological process GO term refers to the biological process to which the gene or gene product contributes. A molecular function GO term is defined as the biochemical activity of a gene product. The third category, cellular component, describes the cellular location where a gene product is active (Ashburner et al., 2000). GO terms are organised so that higher or parent terms describe a broad function, for example cell growth, but lower child terms are more specific, such as pyrimidine biosynthesis. All the genes annotated with a child GO term are also included within the broad parent term. Thus a balance exists between identifying from microarray data, a large number of genes involved with a broad GO term function and identifying very few or no genes from a more specific level of GO term.

To facilitate the identification of genes by GO term a programme such as Babelomics can be used. Babelomics is a web based tool for the functional analysis of genome scale experiments, such as microarray experiments (Al-Shahrour et al., 2006). A variety of tools are available on the Babelomics website (http://babelomics.bioinfo.cipf.es) to help in the functional annotation and analysis of large scale experiments. For example, the FatiGO tool can be used to identify the functional terms of a list of genes and can compare two lists of genes to identify if certain functions have been enriched. For example, the functions of significantly differentially regulated genes identified during a microarray experiment could be compared to the functions of genes unaffected by the experimental parameters. The FatiGO tool can identify which functional terms are over or under represented in the experimental data set to try and elucidate the biological processes affected. Such functional studies are undertaken in an attempt to understand the global response of a biological system rather than the responses of individual genes.

5.1.2 Aims

Only a small portion of the genes identified as differentially regulated in the *7.2sox10*:GFP data set were assessed by *in situ* hybridization. All genes that were validated by *in situ* hybridization as either NC or otic vesicle expressed displayed differential gene expression between WT and *sox10* mutant embryos. As such it would be expected that genes

published as being expressed at these sites would also follow this pattern and display differential gene expression. A number of known NC expressed genes were identified as differentially regulated in the microarray data but were not assessed during the *in situ* hybridization screen as a result of time constraints. A selection of genes that display very strong NC expression in WT embryos are given in **Table 46**. This suggests that there are other genes differentially regulated in *sox10* mutants that have been identified by microarray analysis that await validation. These genes could be identified from the *7.2sox10*:GFP data set. However two other data sets exist, the *4.9sox10*:GFP data set and the *7.2+4.9* combined data set. These data sets could also be examined for interesting targets that are differentially regulated. The aim of this chapter is to assess which is the best microarray data set from which to identify further targets of *sox10*. In addition, methods to interrogate the microarray data sets for genes that are good candidates for validation will be explored. The aim of this chapter therefore is to maximise the value of the microarray data by identifying genes to prioritise for future validation.

Affymetrix ID	Gene Symbol	7.2 *t*-Test *p*-value
Dr.21790.1.A1_at	*pcdh10a*	0.046600606
Dr.17217.1.S1_at	*rab32*	0.032265697
Dr.26358.1.A1_at	*ppat*	0.007318902
Dr.4801.1.S1_at	*gart*	0.006213317

Table 46: Examples of genes identified as differentially expressed but not assessed by *in situ* hybridization.
All genes have been identified as NC expressed on the ZFIN website (Thisse et al., 2004).

5.2 Results and Conclusions

5.2.1 Analysing the Microarray Data Sets

Three microarray data sets have been produced during this project. Two data sets were based on RNA from cells purified from two different transgenic lines. The *7.2sox10*:GFP data set was comprised of three WT arrays versus three mutant arrays. These arrays were hybridised with RNA from NC and otic epithelium GFP positive cells from the *7.2sox10*:GFP transgenic zebrafish line. The *4.9sox10*:GFP data set was comprised of two WT arrays versus two mutant arrays. These arrays were hybridised with RNA from the *4.9sox10*:GFP transgenic line which labels NCCs strongly with GFP but otic epithelium cells only weakly. As both transgenic lines strongly marked NCCs, the two sets of arrays were combined for an analysis of five WT arrays versus five *sox10* mutant arrays. This section aims to assess which of these data sets could be used most reliably to cherry pick interesting targets for validation. This will be performed by analysing the performance of known NC and otic vesicle marker genes expressed at approximately 24 hpf, including those genes validated in Chapter 4, within the three data sets.

5.2.1.1 Performance of NC expressed genes

The *7.2sox10*:GFP data set was analysed for genes significantly differentially regulated using a *t*-Test. The genes were selected for validation by *in situ* hybridization based on ranking by *t*-statistic. This led to 22 NC expressed genes being identified as down-regulated in *sox10* mutant embryos. These genes were represented by 25 probe sets in the 100 most significantly down-regulated genes of the *7.2sox10*:GFP data set. To assess whether these validated genes were also identified in the other data sets, the results (*t*-Test derived *p*-value) were examined for these validated probe sets across all three microarray data sets (**Table 47**). Ten of these validated probe sets were not identified as down-regulated in the *4.9sox10*:GFP data set. A single validated probe set for *zgc:110239* was not identified as down-regulated in the *7.2+4.9* combined data set (**Table 47**). Thus, in contrast to the combined data set, the *4.9sox10*:GFP data set failed to identify many of these validated genes as differentially regulated. To assess the performance of the three microarray data sets without using results biased towards the *7.2sox10*:GFP data set, a selection of well known NC marker genes were examined (**Table 48**). These 10 genes were represented by 11 separate probe sets on the zebrafish Affymetrix array as *gch* is represented by two probe sets. All these genes are known to be expressed in cells of the NC at approximately 24 hpf (Thisse et al., 2004). As all NC expressed genes evaluated during the *in situ* hybridization screen were validated, these genes are also likely to be down-regulated in *sox10* mutant embryos. Indeed, *crestin, dct, foxd3* and *mitfa* have been examined in *sox10* mutant embryos and shown to be down-regulated (Dutton et al., 2001a, Elworthy et al., 2003, Kelsh et al., 2000, Kelsh and Eisen, 2000, Lopes et al., 2008). The combined data set was the only data set to identify all these NC markers as down-regulated in *sox10* mutant embryos. The *7.2sox10*:GFP data set identified 6 probe sets as down-regulated, surprisingly failing to identify *gch* in this category. The *4.9sox10*:GFP data set only identified 4 probe sets as down-regulated. The combined data set identified genes, such as *kita* and *csf1r*, which were not identified as down-regulated in either of the individual data sets (**Table 48**). The combined data set thus appears to be the most reliable dataset to examine for genes to validate.

Table 47: Summary of all NC expressed probe sets validated by *in situ* hybridization in Chapter 4.

The *t*-Test derived *p*-values are given for all probe sets that correspond to NC expressed genes validated in Chapter 4. Gene names were determined, when available, by interrogating the NetAffx database with Affymetrix identifiers. Significantly down-regulated results are highlighted in green, non-significant results are highlighted in red. (See next page)

Affymetrix ID	Gene Symbol	7.2 *t*-Test *p*-value	4.9 *t*-Test *p*-value	7.2+4.9 *t*-Test *p*-value
Dr.1251.1.S1_at	*degs1*	0.00330296	0.17091808	0.01819284
Dr.7966.1.S1_at	*atp6v1e1*	0.00267154	0.03115208	0.00001058
Dr.7250.1.A1_at	*zgc:110239*	0.00261661	0.62630892	0.30263367
Dr.4751.2.A1_at	*aldh2b*	0.00233465	0.01113994	0.00009979
Dr.10336.1.S1_at	*dct*	0.00220708	0.03181335	0.00005568
Dr.13359.1.S1_at	*adsl*	0.00209284	0.10092666	0.00037379
Dr.20896.1.S1_at	*slc2a15b*	0.00180482	0.02056247	0.00005038
Dr.4171.1.A1_at	*pah*	0.00173763	0.00474269	0.00002168
Dr.10624.2.S1_a_at	*zgc:110343*	0.00152714	0.00239700	0.00003745
Dr.8124.1.S1_at	*crestin*	0.00150991	0.08672696	0.00187139
Dr.4751.1.S1_a_at	*aldh2b*	0.00149498	0.03365096	0.00002157
Dr.983.1.S1_at	*paics*	0.00148799	0.02142708	0.00003690
Dr.8594.1.S1_at	*zgc:100919*	0.00135927	0.04615977	0.00000230
Dr.8080.1.S1_at	*mitfa*	0.00130524	0.06132047	0.00018682
Dr.10624.2.S1_at	*zgc:110343*	0.00119537	0.00416982	0.00016658
Dr.3972.1.S1_at		0.00114965	0.00889788	0.00000121
Dr.590.1.S1_at	*foxd3*	0.00103716	0.08952936	0.01265310
Dr.481.1.S1_at	*atic*	0.00097499	0.09128246	0.00004416
DrAffx.1.65.S1_at	*cx33.8*	0.00096602	0.06422850	0.00053927
Dr.2901.1.S1_at	*coro1c*	0.00073671	0.19995388	0.00701491
Dr.14747.1.A1_at	*slc2a15b*	0.00056468	0.04337164	0.00001950
Dr.23456.1.S1_at	*pax7*	0.00044960	0.66678756	0.00174165
Dr.10292.1.S1_at	*rbp4l*	0.00008840	0.00056999	0.00000138
Dr.2855.1.A1_a_at	*atp6v1a*	0.00008356	0.00043360	0.00000078
Dr.4612.1.A1_at	*wu:fc31e04*	0.00005447	0.01980770	0.00000783

Affymetrix ID	Gene Symbol	7.2 *t*-Test *p*-value	4.9 *t*-Test *p*-value	7.2+4.9 *t*-Test *p*-value
Dr.8096.1.S1_at	*kita*	0.26101655	0.07901513	0.02429431
Dr.14668.2.S1_at	*gch*	0.12587915	0.03146349	0.00446180
Dr.10333.1.S1_at	*csf1r*	0.12376060	0.20075732	0.00621486
Dr.14668.1.S1_at	*gch*	0.09166294	0.02026650	0.00329152
Dr.19561.1.S1_at	*tyr*	0.07561973	0.14350338	0.01024735
Dr.9225.1.A1_at	*aox3 (xdh)*	0.01398500	0.00162993	0.00102245
Dr.10456.1.S1_at	*ednrb1*	0.00546011	0.11667577	0.00014209
Dr.10336.1.S1_at	*dct*	0.00220708	0.03181335	0.00005568
Dr.8124.1.S1_at	*crestin*	0.00150991	0.08672696	0.00187139
Dr.8080.1.S1_at	*mitfa*	0.00130524	0.06132047	0.00018682
Dr.590.1.S1_at	*foxd3*	0.00103716	0.08952936	0.01265310

Table 48: Table of microarray data for probe sets of known NC marker genes.
The *t*-Test derived *p*-values are given for probe sets representing known marker genes of the NC. Significantly down-regulated results are highlighted in green, non-significant results are highlighted in red. All these genes are expressed in NCCs at approximately 24 hpf (Thisse et al., 2004).

This study was extended to include all genes on the ZFIN website that have been published as NC expressed (**Table 49**). NC expressed genes (including substructures) were selected by interrogating the anatomy section on the ZFIN website. The list of genes was then restricted to an appropriate time period, 22 to 30 hpf. This gave 130 expressed genes of which 92 had at least one corresponding Affymetrix probe set. These genes were represented by a total of 121 probe sets. This list of genes included genes expressed in skeletogenic NCCs and was therefore not restricted to NCCs dependent on *sox10* expression for specification. The *7.2sox10*:GFP data set identified 47 probe sets as differentially regulated, the *4.9sox10*:GFP data set identified 29 probe sets and the combined data set identified 55 probe sets (**Table 49**). This indicated that the combined data set was the most promising data set to interrogate for genes of interest. Interestingly, while most genes were down-regulated, several NC expressed genes were identified as up-regulated. These genes are tempting candidates for *in situ* hybridization validation and may facilitate the identification of up-regulated NC genes in *sox10* mutant embryos. Indeed all NC expressed genes on ZFIN that have been identified as differentially regulated in the microarray data are targets for validation.

To refine this list of genes to *sox10* dependent NCCs the same search was performed using the term “pigment cell” (**Table 50**). This list of genes will include those expressed in NC derived pigment cells but also those expressed in the melanocytes of the pigmented retinal epithelium, for example *mitfb* (Lister et al., 2001). 50 genes were expressed in pigment cells during the time period of interest. 40 probe sets on the array corresponded to 31 of these genes. The *7.2sox10*:GFP data set identified 26 probe sets as differentially regulated, the *4.9sox10*:GFP data set identified 15 probe sets and the combined data set identified 27 probe sets. Thus similar numbers of pigment cell expressed genes were identified in the *7.2sox10*:GFP and combined data sets with the *4.9sox10*:GFP data set performing less well. It is interesting to note that of the 31 pigment cell genes represented on the Affymetrix array, 26 were identified as down-regulated in one or more of the microarray data sets. A large number of melanocyte or xanthophore expressed genes were validated by *in situ* hybridization in Chapter 4. These two observations both suggest that these sets of microarray data are powerful tools for identifying pigment cell expressed genes. In summary, the combined data set appears to be a more reliable indicator of differential gene regulation of NC expressed genes in *sox10* mutant embryos. This could be a result of the increased sample size examined in this data set. To definitively test this, a blind *in situ* hybridization validation screen would need to be performed to examine the hit rate of NC and otic vesicle expressed genes compared to the *7.2sox10*:GFP results presented in Chapter 4.

The only PNS expressed gene successfully validated during the *in situ* hybridization screen was *foxd3*. This is consistent with the fact that there are very few

markers available for the PNS at 24 hpf. In an attempt to reveal PNS markers and to provide candidate genes to test for differential PNS gene expression in *sox10* mutant embryos, ZFIN anatomy searches were performed as above. The anatomical search terms used were “neuron, neural crest derived” (**Table 51**) and “glial cell” (**Table 52**). On ZFIN, 79 genes have been identified as expressed in NC derived neurons and of these 27 were expressed between 22 and 30 hpf. 16 of these genes were represented by 18 Affymetrix probe sets. 6 of these genes were identified as up-regulated in one of the three microarray data sets, no genes were identified as down-regulated and no genes were identified as up-regulated in two or more data sets (**Table 51**). This is in stark contrast to the results for the pigment cell expressed genes, all of which were down-regulated but more importantly, all of which were identified as such in at least two of the data sets. Thus the results for these neuronal expressed genes may not be reliable. However, as only a few neuronal expressed genes are known to be differentially expressed in *sox10* mutants, these genes are worth examining by *in situ* hybridization. 28 genes have been annotated as glial cell expressed on ZFIN, of these only 3 were expressed during the time period of interest (**Table 52**). From these genes, *foxd3* was identified as down-regulated in two data sets and has been validated as down-regulated in glia previously (Kelsh et al., 2000). Neither of the other two genes were identified as differentially regulated. *Mbp* is not expressed at 24 hpf (Thisse et al., 2004) thus it is not surprising that this gene was not identified as differentially regulated in this microarray screen. Thus this approach has not helped to uncover more glial expressed genes that are likely to be down-regulated in *sox10* mutant embryos.

Affymetrix ID	Gene Symbol	7.2 *t*-Test *p*-value	4.9 *t*-Test *p*-value	7.2+4.9 *t*-Test *p*-value	
No	*acbd7*				
Dr.20912.1.S1_at	*alcam*	0.44367251	0.18985467	0.65715164	
Dr.20912.1.S2_at	*alcam*	0.94103247	0.18985467	0.65715164	
Dr.4751.1.S1_a_at	*aldh2b*	0.00149498	0.03365096	0.00002157	Down
Dr.4751.2.A1_at	*aldh2b*	0.00233465	0.01113994	0.00009979	Down
Dr.14515.1.A1_at	*aldh16a1*	0.06206876	0.40363121	0.02175599	Down
Dr.9225.1.A1_at	*aox3 (xdh)*	0.01398500	0.00162993	0.00102245	Down
Dr.24131.1.S1_at	*ap3b1*	0.03986506	0.32664850	0.02191994	Down
Dr.24201.1.A1_at	*ap3b1*	0.12334988	0.29018345	0.09773549	
Dr.481.1.S1_at	*atic*	0.00097499	0.09128246	0.00004416	Down
Dr.19482.1.A1_at	*barx1*	0.02183935	0.76380199	0.26302573	Down
Dr.12383.2.S1_a_at	*bcmo1*	0.75866866	0.14405601	0.20674919	
Dr.12376.1.A1_at	*bmper*	0.76941806	0.21464749	0.69242579	
No	*c10orf11*				
Dr.14422.1.A1_at	*cdon*	0.06976867	0.67238581	0.26811984	
Dr.14422.3.A1_x_at	*cdon*	0.04962551	0.02272565	0.00605113	UP
Dr.14422.2.S1_at	*cdon*	0.00608916	0.05217386	0.00024247	UP
Dr.6570.1.S1_at	*cited3*	0.09305646	0.48906589	0.42471692	
Dr.6570.1.S2_at	*cited3*	0.18408886	0.50958717	0.26852319	

Dr.6571.1.S1_at	*cndp2*	0.07980923	0.10165119	0.05346033	
Dr.8124.1.S1_at	*crestin*	0.00150991	0.08672696	0.00187139	Down
No	*crestinl*				
Dr.10333.1.S1_at	*csf1r*	0.12376060	0.20075732	0.00621486	Down
Dr.3374.2.S1_at	*ctsba*	0.04551315	0.14553307	0.37332499	Down
Dr.4782.1.S1_at	*ctsc*	0.01052202	0.17543989	0.00098793	Down
Dr.5040.1.S1_at	*cyb5a*	0.02398859	0.03205880	0.00046595	Down
Dr.1251.1.S1_at	*degs1*	0.00330296	0.17091808	0.01819284	Down
Dr.19211.1.S1_at	*dera*	0.03396077	0.13432646	0.01961480	Down
Dr.21797.1.S1_at	*dfna5*	0.47294620	0.42051473	0.68788755	
Dr.7949.1.S1_at	*dlg1*	0.12859283	0.73419821	0.30322888	
Dr.2.1.S1_at	*dlx2a*	0.08324626	0.18918946	0.10120377	
Dr.17477.1.S1_at	*dtnbp1a*	0.06309766	0.12429862	0.03482805	Down
Dr.12576.1.S1_at	*ednra*	0.51036042	0.30257273	0.78283429	
Dr.20789.1.A1_at	*erbb2*	0.28649518	0.14815147	0.03358486	Down
Dr.12366.1.S1_at	*ets1a*	0.92823583	0.56444848	0.99816859	
Dr.12366.2.A1_at	*ets1a*	0.19307689	0.43550211	0.12932420	
No	*evi1*				
Dr.3713.1.A1_at	*flncb*	0.25183645	0.66238552	0.45772052	
Dr.590.1.S1_at	*foxd3*	0.00103716	0.08952936	0.01265310	Down
Dr.13161.1.S1_at	*gadd45bl*	0.43283209	0.47833639	0.14824764	
Dr.4801.1.S1_at	*gart*	0.00621332	0.00762501	0.00086417	Down
Dr.14668.1.S1_at	*gch2*	0.09166294	0.02026650	0.00329152	Down
Dr.14668.2.S1_at	*gch2*	0.12587915	0.03146349	0.00446180	Down
Dr.3275.1.A1_at	*gpd1b*	0.03814965	0.12779483	0.00584716	Down
No	*gpr143*				
Dr.23788.1.S1_at	*gstp1*	0.01244999	0.08942329	0.04036518	Down
Dr.6382.1.A1_at	*hhip*	0.11818337	0.07007825	0.01270617	UP
Dr.5772.1.S1_at	*hoxa2b*	0.17399095	0.40500796	0.65684378	
Dr.15717.1.S1_at	*hoxb1a*	0.02427962	0.12227093	0.77110028	UP
Dr.20962.1.S1_at	*hoxb2a*	0.21660677	0.88262790	0.83758640	
Dr.10626.1.A1_at	*hoxb2a*	0.55855376	0.29411480	0.08929119	
Dr.5779.1.S1_at	*hoxb3a*	0.36674219	0.01520381	0.07713863	UP
Dr.5428.1.S1_at	*id2a*	0.16466472	0.38644043	0.18560024	
Dr.12583.1.S1_at	*igfbp3*	0.67537624	0.94783461	0.43605354	
Dr.12583.2.A1_at	*igfbp3*	0.52644509	0.45061094	0.85839808	
Dr.12379.1.S1_at	*il17rd*	0.07022136	0.17281879	0.02638596	UP
No	*im:7140212*				
No	*im:7142864*				
No	*si:ch211-222e23.2*				
Dr.2855.2.S1_at	*atp6v1a*	0.12184907	0.30382824	0.07631407	
Dr.2855.1.A1_a_at	*atp6v1a*	0.00008356	0.00043360	0.00000078	Down
Dr.8096.1.S1_at	*kita*	0.26101655	0.07901513	0.02429431	Down
No	*lratb*				
No	*ltk*				
Dr.608.1.A1_at	*mafba*	0.18024163	0.72081417	0.48328298	
Dr.1680.1.S1_at	*meis3*	0.04055930	0.04795788	0.00354967	UP
Dr.8080.1.S1_at	*mitfa*	0.00130524	0.06132047	0.00018682	Down
Dr.2596.2.A1_a_at	*mpv17*	0.11796077	0.10906474	0.01362077	UP
Dr.2596.2.A1_x_at	*mpv17*	0.14466071	0.04209826	0.00810471	UP
Dr.2596.5.A1_x_at	*mpv17*	0.37744278	0.41452524	0.05553884	
Dr.17843.1.A1_at	*mpv17*	0.03258898	0.42932430	0.91304076	Down
Dr.25316.1.S1_at	*nlcam*	0.26748446	0.81812811	0.40730393	
Dr.4171.1.A1_at	*pah*	0.00173763	0.00474269	0.00002168	Down
Dr.983.1.S1_at	*paics*	0.00148799	0.02142708	0.00003690	Down

Dr.579.1.S1_at	*pax3a*	0.42822087	0.22187571	0.07897566	
No	*pax3b*				
Dr.23456.1.S1_at	*pax7a (pax7e)*	0.00044960	0.66678756	0.00174165	Down
Dr.8033.1.S1_at	*pax7a (pax7d)*	0.03809424	0.38300705	0.12114859	Down
Dr.23455.1.S1_at	*pax7a (pax7c)*	0.20390600	0.25540304	0.21863870	
Dr.1470.1.S1_at	*pax7a*	0.82161093	0.92229867	0.95834744	
No	*pax7b*				
Dr.21026.1.S1_at	*pcdh10a*	0.59820735	0.06410612	0.70804060	
Dr.21790.1.A1_at	*pcdh10a*	0.04660061	0.25406221	0.00684118	Down
Dr.12.1.S1_at	*pdgfra*	0.57346189	0.71522051	0.42611179	
Dr.12.1.S2_at	*pdgfra*	0.06314114	0.11937556	0.05489326	
Dr.10257.1.A1_at	*pepd*	0.19357376	0.33827218	0.11715376	
No	*plcb3*				
Dr.26392.2.S1_at	*pold1*	0.61816293	0.54821336	0.65618092	
Dr.9175.1.A1_at	*pold1*	0.24453563	0.20382360	0.04444419	UP
Dr.26358.1.A1_at	*ppat*	0.00731890	0.03197975	0.00065635	Down
Dr.14001.1.A1_at	*pttg1ip*	0.01807404	0.02672966	0.00143626	Down
Dr.14001.2.S1_at	*pttg1ip*	0.04133245	0.03335282	0.00275114	Down
Dr.12260.1.A1_at	*rab7*	0.00403438	0.06317651	0.00093235	Down
Dr.17217.1.S1_at	*rab32*	0.03226570	0.01295175	0.00041274	Down
Dr.193.1.S1_at	*raraa*	0.14747715	0.67268211	0.05884251	
Dr.23544.1.S1_s_at	*rarga*	0.83216679	0.37732565	0.67024988	
Dr.194.1.S1_at	*rarga*	0.31954211	0.04755262	0.85664773	Down
No	*rargb*				
Dr.10292.1.S1_at	*rbp4l*	0.00008840	0.00056999	0.00000138	Down
Dr.15169.1.A1_at	*rgs14*	0.90088600	0.10029853	0.16752151	
No	*rxraa*				
Dr.348.1.S1_at	*rxrga*	0.97731149	0.64712524	0.74930751	
No	*sb:cb429*				
No	*sb:eu188*				
No	*sb:eu233*				
No	*sb:eu383*				
No	*sb:eu823*				
No	*sdc4*				
No	*sdc4*				
No	*selj*				
Dr.8061.1.S1_at	*sema3d*	0.01972223	0.06748841	0.00402466	UP
No	*si:busm1-172d23.2*				
Dr.19537.1.S1_a_at	*si:dkey-31f5.7*	0.12623142	0.36014467	0.50056982	
Dr.25685.1.A1_at	*si:dkey-31f5.7*	0.03678070	0.03715282	0.00512837	Down
No	*si:dkey-251i10.2*				
No	*slc39a1*				
No	*slc45a2*				
Dr.625.1.S1_at	*snai1b*	0.93246692	0.32032564	0.40757066	
Dr.20.3.S1_x_at	*snai2*	0.70444691	0.69350541	0.92420554	
Dr.20.1.S1_at	*snai2*	0.04492790	0.04804532	0.01577159	Down
Dr.20.3.S1_at	*snai2*	0.80465686	0.81674427	0.54248405	
DrAffx.2.85.S1_at	*sox4b*	0.11148241	0.02486156	0.98327315	Down
Dr.12604.1.S1_at	*sox10*	0.13209380	0.51710188	0.34151468	
Dr.10038.1.S1_at	*spry1*	0.33454302	0.89178789	0.97663689	
Dr.923.1.A1_at	*sqstm1*	0.10544427	0.11012185	0.00469088	Down
Dr.15842.1.A1_at	*syngr1*	0.02023073	0.00469429	0.00033034	Down
No	*tbc1d5*				
Dr.8506.1.S1_a_at	*tfap2a*	0.93010110	0.47889030	0.98553759	

Dr.9638.1.S1_a_at	*tnc*	0.85379112	0.32953891	0.98698306	
Dr.9638.1.A1_a_at	*tnc*	0.27302995	0.98284644	0.58994442	
Dr.8121.1.S1_at	*twist1a*	0.14644158	0.64330959	0.08623191	
No	*twist1b*				
Dr.8127.1.S1_at	*ube2i*	0.13066502	0.05531247	0.01923041	UP
Dr.389.1.S1_at	*wnt5b*	0.64728123	0.16993344	0.17441466	
Dr.8822.1.A1_at	*zeb2a (sip1a)*	0.00374613	0.14579193	0.00891323	Down
DrAffx.2.73.S1_at	*zeb2a (sip1a)*	0.55840504	0.30754045	0.59923309	
No	*zeb2b*				
No	*zgc:63474*				
Dr.4031.1.A1_at	*zgc:73134*	0.08620743	0.49211428	0.06297540	
Dr.16684.1.S1_at	*zgc:77051*	0.14692816	0.46021384	0.41851234	
Dr.890.1.S1_at	*zgc:77517*	0.07359646	0.95719385	0.17930047	
Dr.16185.1.S1_at	*zgc:77775*	0.00761450	0.11927229	0.00187873	Down
No	*zgc:85942*				
No	*zgc:91909*				
Dr.9944.1.A1_at	*zgc:91968*	0.01952465	0.01008070	0.00096157	Down
No	*zgc:101606*				
No	*zgc:103530*				
No	*zgc:103749*				
Dr.7250.1.A1_at	*zgc:110239*	0.00261661	0.62630892	0.30263367	Down
Dr.10624.2.S1_a_at	*zgc:110343*	0.00152714	0.00239700	0.00003745	Down
Dr.10624.2.S1_at	*zgc:110343*	0.00119537	0.00416982	0.00016658	Down
Dr.10624.1.S1_at	*zgc:110343*	0.00851871	0.00023531	0.00020797	Down
Dr.22923.1.A1_at	*zgc:110355*	0.00404091	0.09509496	0.00240261	Down
Dr.5081.1.A1_at	*zgc:110784*	0.05444692	0.24116734	0.05048341	
Dr.12707.1.A1_at	*zgc:112072*	0.00097298	0.04124477	0.00017576	Down
No	*zgc:112101*				
No	*zgc:113337*				
No	*zgc:113456*				
Dr.10102.2.S1_at	*zgc:136369*	0.45216620	0.60071957	0.38222027	
Dr.10102.1.A1_at	*zgc:136369*	0.91889745	0.19289894	0.60223079	
No	*zgc:162334*				

Table 49: Microarray results for ZFIN NC expressed genes.
A ZFIN anatomy search was performed using the term neural crest. All NC genes expressed between the 20-25 somite stage and Prim 15 stage (approximately 22 hpf to 30 hpf) were selected and are presented alphabetically. Affymetrix IDs were determined from the NetAffx website and *t*-Test *p*-values are given for each microarray data set. The far right column indicates whether a probe set was up or down-regulated. (See previous page)

Affymetrix ID	Gene Symbol	7.2 *t*-Test *p*-value	4.9 *t*-Test *p*-value	7.2+4.9 *t*-Test *p*-value
Dr.4751.1.S1_a_at	*aldh2b*	0.0014950	0.0336510	0.0000216
Dr.4751.2.A1_at	*aldh2b*	0.0023346	0.0111399	0.0000998
Dr.9225.1.A1_at	*aox3*	0.0139850	0.0016299	0.0010225
Dr.15693.1.S1_at	*atp6v0c*	0.1971969	0.3194535	0.1264327
Dr.1971.1.A1_at	*atp6v0c*	0.0093238	0.1356747	0.0002904
Dr.3514.1.S1_at	*cox4nb*	0.0065779	0.7956251	0.1957595
Dr.10333.1.S1_at	*csf1r*	0.1237606	0.2007573	0.0062149
Dr.10336.1.S1_at	*dct*	0.0022071	0.0318134	0.0000557
Dr.19211.1.S1_at	*dera*	0.0339608	0.1343265	0.0196148
Dr.17477.1.S1_at	*dtnbp1a*	0.0630977	0.1242986	0.0348281

Dr.10456.1.S1_at	*ednrb1*	0.0054601	0.1166758	0.0001421
Dr.3713.1.A1_at	*flncb*	0.2518364	0.6623855	0.4577205
Dr.4801.1.S1_at	*gart*	0.0062133	0.0076250	0.0008642
Dr.14668.1.S1_at	*gch2*	0.0916629	0.0202665	0.0032915
Dr.14668.2.S1_at	*gch2*	0.1258792	0.0314635	0.0044618
No	*gmps*			
No	*gpr143*			
No	*gpr182*			
Dr.23788.1.S1_at	*gstp1*	0.0124500	0.0894233	0.0403652
Dr.14058.1.A1_at	*gstt1a*	0.9572253	0.3944546	0.6610474
Dr.25191.1.S1_at	*idh1*	0.0300042	0.6534362	0.4241687
Dr.14555.1.S1_at	*ivns1abpa*	0.0387004	0.4871136	0.0420197
Dr.9512.1.A1_at	*ivns1abpa*	0.0076884	0.1765334	0.0005903
Dr.8096.1.S1_at	*kita*	0.2610165	0.0790151	0.0242943
Dr.729.1.S1_at	*ldhb*	0.0222297	0.1858931	0.1338363
No	*ltk*			
Dr.8080.1.S1_at	*mitfa*	0.0013052	0.0613205	0.0001868
Dr.12631.1.S1_at	*mitfb*	0.3067101	0.4362738	0.3427937
Dr.24311.1.S1_at	*nme4*	0.0451185	0.1693777	0.0442447
Dr.4171.1.A1_at	*pah*	0.0017376	0.0047427	0.0000217
Dr.23456.1.S1_at	*pax7a (pax7e)*	0.0004496	0.6667876	0.0017417
Dr.8033.1.S1_at	*pax7a (pax7d)*	0.0380942	0.3830070	0.1211486
Dr.23455.1.S1_at	*pax7a (pax7c)*	0.2039060	0.2554030	0.2186387
Dr.1470.1.S1_at	*pax7a*	0.8216109	0.9222987	0.9583474
No	*pax7b*			
Dr.11425.1.S1_at	*psat1*	0.4161273	0.8670802	0.5461265
Dr.14001.1.A1_at	*pttg1ip*	0.0180740	0.0267297	0.0014363
Dr.14001.2.S1_at	*pttg1ip*	0.0413324	0.0333528	0.0027511
Dr.17217.1.S1_at	*rab32*	0.0322657	0.0129518	0.0004127
No	*sb:eu188*			
Dr.19537.1.S1_a_at	*si:dkey-31f5.7*	0.1262314	0.3601447	0.5005698
Dr.25685.1.A1_at	*si:dkey-31f5.7*	0.0367807	0.0371528	0.0051284
No	*silva*			
No	*slc24a5*			
No	*slc39a10*			
No	*slc45a2*			
No	*spra*			
Dr.15842.1.A1_at	*syngr1*	0.0202307	0.0046943	0.0003303
No	*tyrp1b*			
Dr.1368.4.A1_at	*wu:cegs2586*	0.9591529	0.6867187	0.6530809
No	*zgc:77375*			
No	*zgc:86889*			
Dr.9944.1.A1_at	*zgc:91968*	0.0195247	0.0100807	0.0009616
Dr.8594	*zgc:100919*	0.0013593	0.0461598	0.0000023
No	*zgc:103530*			
No	*zgc:110789*			
No	*zgc:162334*			
No	*zgc:162606*			

Table 50: Microarray results for ZFIN pigment cell expressed genes.
A ZFIN anatomy search was performed using the term pigment cell as performed for Table 4. All differentially regulated genes were down-regulated.

Affymetrix ID	Gene Name	7.2 *t*-Test *p*-value	4.9 *t*-Test *p*-value	7.2+4.9 *t*-Test *p*-value
Dr.15722.1.S1_at	*ache*	0.408349216	0.636530876	0.865899742
Dr.20912.1.S1_at	*alcam*	0.443672508	0.281618953	0.998932274
Dr.20912.1.S2_at	*alcam*	0.941032469	0.189854667	0.65715164
No	*chrna4*			
Dr.6798.1.S1_at	*cxcr4b*	0.047582835	0.985801399	0.411663681
Dr.10492.1.S1_at	*dla*	0.092713371	0.032959372	0.212765485
No	*flot1a*			
Dr.17863.1.S1_at	*flot2a*	0.723339081	0.422459006	0.808896005
Dr.15717.1.S1_at	*hoxb1a*	0.024279619	0.122270934	0.771100283
No	*insm1a*			
Dr.19235.1.S1_at	*insm1b*	0.047903273	0.73179245	0.126329988
Dr.12208.1.S1_s_at	*isl1*	0.355698913	0.430072218	0.684007704
Dr.279.1.S1_at	*isl1*	0.670866251	0.844222367	0.587054551
Dr.281.1.S1_at	*isl2a*	0.349266678	0.553344309	0.191186607
Dr.570.1.S1_at	*lhx3*	0.095309049	0.890643537	0.337820679
No	*met*			
No	*mnx1*			
No	*nkx6.1*			
Dr.26440.1.S1_at	*nrp1a*	0.047821794	0.476297349	0.294988781
Dr.10562.1.A1_at	*olfm2*	0.341941446	0.551333189	0.439760268
Dr.3920.1.S1_at	*olig2*	0.821317732	0.486432523	0.481608063
Dr.24244.2.S1_a_at	*pax6a*	0.15213044	0.947506189	0.263120949
No	*phox2a*			
	plxna3			
Dr.439.1.A1_at	*prickle1b*	0.312992334	0.184240609	0.033645876
Dr.12382.1.S1_at	*slc6a3*	0.316994131	0.982832909	0.080466352
No	*sp8*			

Table 51: Microarray results for ZFIN NC derived neuron expressed genes.
A ZFIN anatomy search was performed using the term neuron NC derived, as performed for Table 4. All differentially regulated genes were up-regulated.

Affymetrix ID	Gene Name	7.2 *t*-Test *p*-value	4.9 *t*-Test *p*-value	7.2+4.9 *t*-Test *p*-value
Dr.590.1.S1_at	*foxd3*	0.001037158	0.089529358	0.012653097
Dr.7462.1.A1_at	*mbp*	0.276803732	0.824697912	0.501458645
Dr.3920.1.S1_at	*olig2*	0.821317732	0.486432523	0.481608063

Table 52: Microarray results for ZFIN glial cell expressed genes.
A ZFIN anatomy search was performed using the term glial cell as performed for Table 4. *Foxd3* was identified as down-regulated.

5.2.1.2 Performance of otic vesicle expressed genes

Five otic vesicle expressed genes were identified as differentially regulated in *sox10* mutant embryos from the *7.2sox10*:GFP data and were subsequently validated. Results from the microarray analysis (*t*-Test derived *p*-value) were examined for these genes across all three microarray data sets (**Table 53**). It was expected that the otic vesicle expressed genes would not be readily identified as differentially regulated in the *4.9sox10*:GFP data set because otic epithelium cells are only weakly labelled with GFP in

this line. This appeared to be true as only *otomp* was identified as differentially regulated in the *4.9sox10*:GFP line. Combining the two data sets also appeared to reduce the power of the screen when compared to the *7.2sox10*:GFP data. Only two of the genes validated in Chapter 4 were identified as differentially regulated in the combined data set (**Table 53**). However, a paucity of validated otic vesicle markers known to be differentially regulated in *sox10* mutant embryos makes this analysis difficult. Recently a study examining the otic vesicle phenotype of *sox10* mutant embryos was conducted (Dutton et al., 2009). The expression patterns of some otic vesicle marker genes were examined in WT and *sox10* mutant embryos. The genes which were up-regulated in *sox10* mutant embryos at 24 hpf are shown in **Table 54**; no genes were down-regulated. These up-regulated genes were poorly detected in the three microarray data sets with only a few probe sets being identified as differentially regulated in one of the three microarray data sets. Two of these differentially regulated probe sets corresponded to *sox9b* and these showed conflicting results in the *4.9sox10*:GFP data set, one being called up-regulated and one, against the published data, being called down-regulated (**Table 54**). Thus the results were not consistent across the three data sets and with the published data. Indeed it appears that otic vesicle expressed genes were not easily detected as differentially-regulated in any data set.

Affymetrix ID	Gene Symbol	7.2 *t*-Test *p*-value	4.9 *t*-Test *p*-value	7.2+4.9 *t*-Test *p*-value	
Dr.21012.1.A1_at	*sfrp5*	0.00292677	0.89689237	0.27037165	Up
Dr.23461.1.A1_at	*stc2*	0.00215093	0.25583288	0.00428666	Up
Dr.19416.1.S1_at	*otomp*	0.00267353	0.01195940	0.01608922	Down
Dr.18158.1.A1_at		0.00152277	0.12411628	0.18378794	Down
Dr.8680.1.S1_at	*cldnj*	0.00008655	0.13460585	0.07455172	Down

Table 53: Summary of all otic vesicle expressed genes validated in Chapter 4.
The *t*-Test derived *p*-values are given for all probe sets that correspond to otic vesicle expressed genes validated in Chapter 4. Significantly differentially regulated results are highlighted in green, non-significant results are highlighted in red. Direction of change, up or down-regulation, is indicated in the far right column.

Affymetrix ID	Gene Symbol	7.2 *t*-Test *p*-value	4.9 *t*-Test *p*-value	7.2+4.9 *t*-Test *p*-value	
Dr.198.1.S1_at	*fsta*	0.917976379	0.988044858	0.848030746	
Dr.198.1.S2_at	*fsta*	0.079166167	0.090070494	0.023546828	Up
Dr.567.2.S1_a_at	*bmp4*	0.738948643	0.284754723	0.831891239	
Dr.478.1.S1_at	*fgf8a*	0.492221892	0.652610123	0.384049505	
Dr.11850.1.S2_at	*sox9b*	0.260425121	0.024585784	0.020765815	Up
Dr.11850.1.S1_at	*sox9b*	0.827677131	0.047364879	0.218093172	Down
Dr.10460.1.S1_at	*sox9a*	0.077197567	0.906565189	0.18403475	
Dr.385.1.S1_at	*wnt4a*	0.035770889	0.699507535	0.30663532	Up

Table 54: Summary of otic vesicle expressed genes known to be up-regulated in *sox10* mutant embryos.
The *t*-Test derived *p*-values are given for probe sets representing genes known to be up-regulated in *sox10* mutant embryos at approximately 24 hpf (Dutton et al., 2009). Significantly differentially regulated results are highlighted in green, non-significant results are highlighted in red. Direction of change, up or down-regulation, is indicated in the far right column.

5.2.2 Identifying Good Candidates for Validation

The aim of this section is to identify genes that are good candidates for further validation, including genes with no prior association to NC or otic vesicle biology. This will be done by applying new approaches to evaluate the existing microarray data thus generating gene lists of interesting candidates. These lists will be examined to identify known NC and otic vesicle expressed genes, known false positive genes and genes with no known expression pattern by integrating published expression data and the results obtained in Chapter 4. This information will be used to evaluate the success of each method to enable the selection of some high priority genes for future validation. Primarily the combined *7.2+4.9* data set will be assessed for high priority genes as this data set appears to most reliably identify known differentially regulated genes.

5.2.2.1 Identifying Candidates with a large fold change

The combined *7.2+4.9* data set was ranked by fold change values generated from the GEPAS CLEAR test. The genes with the most negative fold change (down-regulated) (**Table 55**) and largest positive fold change (up-regulated) (**Table 56**) were selected using a fold change of 2 (> 2 or < -2) as an arbitrary cut off.

For the down-regulated genes (**Table 55**), 55 probe sets corresponding to 52 genes, had a fold change of < -2. Of these, 24 probe sets already had a strong expression pattern on the ZFIN website including 17 probe sets which displayed NC (13 probe sets) or otic vesicle (4 probe sets) expression. There were also 13 genes in this list that were

validated as NC (12 probe sets) or otic vesicle (1 probe set) expressed and down-regulated in *sox10* mutants during the *in situ* hybridization screen. In the list of genes shown in **Table 55**, 30 probe sets were examined during the *in situ* hybridization screen and as 13 were validated, 17 (57 %) of the probe sets were false positives. When the information from ZFIN and the *in situ* hybridization screen was combined, 20 of the 55 identified probe sets were known to be NC or otic vesicle expressed. That a large number of known NC and otic vesicle expressed genes were present in this gene list increases confidence that unknown genes expressed in the cell types of interest are also present. The genes with a well characterised expression pattern that do not show expression in the NC or otic vesicle are unlikely to be of interest.

The published data could be used to prioritise candidates for validation and to reduce the number of candidates for validation. Caution must be exercised with this technique, for example while *slc3a2* was not annotated as being NC expressed (Thisse et al., 2004) the pattern on ZFIN does show some potential migratory NCCs and this gene is expressed in the dorsal neural tube (DNT), the site of NCCs prior to delamination. The use of published data can also highlight targets examined during the *in situ* hybridization screen that might be worth revisiting, for example *zgc:112072* has been annotated as NC expressed on ZFIN but this was not observed during the screen (Thisse et al., 2004). Genes evaluated during the screen should only be revisited when there is a compelling reason, such as an expression pattern of interest, for doing so. There are 6 genes in **Table 55** which have a poorly characterised expression pattern on ZFIN, described as not spatially restricted. This included the gene *slc2a15b* which was validated during the *in situ* hybridization screen as NC expressed and down-regulated in *sox10* mutant embryos. Thus genes annotated in this way may be expressed in cell types of interest and could be examined. Throughout this chapter such annotated unclear expression patterns will be counted as lacking a defined expression pattern. There are 24 Affymetrix probe sets that correspond to targets that have no published gene expression data within this gene list. Some of these genes were analysed during the *in situ* hybridization screen, for example Dr.3972.1.S1_at was expressed in the neural crest while for some of these genes no NC or otic vesicle expression was observed (false positives). There remain 14 genes that lacked an expression pattern on ZFIN and that were not examined during the *in situ* hybridization screen. This group of genes could be prioritised for *in situ* hybridization validation to facilitate the discovery of novel NC or otic vesicle expressed genes. In summary, genes annotated as NC or otic vesicle expressed on ZFIN that were not examined during the *in situ* hybridization screen should be high priorities for future validation with genes without a published expression pattern prioritised next for examination. Genes with an unclear published expression pattern (not spatially restricted) should then be prioritised for validation. Finally genes with a well defined expression

pattern from published data that is not of interest should have the lowest priority for future validation studies. This order of priority should remain the same regardless of the method employed to select genes for validation in this chapter. Genes within each category should be prioritised by the fold change ranking displayed in **Table 55**.

A large number of NC and otic vesicle marker genes were present in the down-regulated genes that displayed large fold changes. This indicated that the same analysis could unearth up-regulated targets for validation. For the up-regulated genes, 23 probe sets had a fold change greater than 2. 13 of these genes had a published expression pattern on ZFIN, two of which were expressed in the NC. Neither of these genes were examined during the *in situ* hybridization screen, analysis of the expression pattern of *mpv17* and *sema3d* should be a priority to facilitate the identification of up-regulated NC expressed genes. 4 of the genes in this list were examined during the *in situ* hybridization screen, all 4 were false positives. 10 up-regulated genes with large fold changes have no published expression pattern and this group of genes may contain some novel NC or otic vesicle expressed genes. However, in line with the up-regulated gene *in situ* hybridization screen, detection of up-regulated genes may be rare.

Affymetrix ID	Gene Symbol	Fold Change	ZFIN Expression Pattern	ISH Screen
Dr.10336.1.S1_at	*dct*	-14.58	Yes, NC	Yes, NC
Dr.10292.1.S1_at	*rbp4l*	-11.03	Yes, NC	Yes, NC
Dr.25257.1.A1_at	*bhlhb3l*	-8.42		Yes
Dr.518.1.A1_at	*sb:cb319*	-3.88	Yes	Yes
Dr.13919.1.S1_at		-3.85		Yes
Dr.2268.1.A1_at	*wu:fc64c05*	-3.49	No expression pattern	
Dr.14747.1.A1_at	*slc2a15b*	-3.46	Yes, not spatially restricted	Yes, NC
Dr.4612.1.A1_at	*wu:fc31e04*	-3.41	No expression pattern	Yes, NC
Dr.3933.1.A1_at	*wu:fc46b01*	-3.40	No expression pattern	Yes
Dr.3972.1.S1_at		-3.36		Yes, NC
Dr.22614.1.A1_at		-3.35		
Dr.26101.1.A1_at		-3.32		Yes
Dr.4171.1.A1_at	*pah*	-3.30	Yes, NC	Yes, NC
Dr.10624.2.S1_a_at	*zgc:110343*	-3.18	Yes, NC	Yes, NC
Dr.4751.2.A1_at	*aldh2b*	-3.15	Yes, NC	Yes, NC
Dr.14787.1.A1_at	*s100b*	-2.99		
Dr.6142.1.A1_at	*si:dkey-180p18.9*	-2.90	Yes, not spatially restricted	
Dr.2788.1.S1_at	*si:dkeyp-86b9.2*	-2.72	Yes, not spatially restricted	Yes
Dr.22360.1.A1_at	*zgc:85890*	-2.71	Yes	
Dr.20850.1.S1_at	*fabp7a*	-2.69	Yes	Yes
Dr.18194.1.A1_at	*pik3c3*	-2.67	Yes, not spatially restricted	Yes
Dr.4838.1.A1_at	*zgc:158673*	-2.56	Yes	Yes
Dr.8124.1.S1_at	*crestin*	-2.53	Yes, NC	Yes, NC
Dr.10624.2.S1_at	*zgc:110343*	-2.48	Yes, NC	Yes, NC
Dr.3991.1.A1_at	*wu:fd02h10*	-2.45	No expression pattern	Yes
Dr.11035.1.A1_at	*wu:fj35f12*	-2.44	No expression pattern	
Dr.9896.1.A1_at	*si:ch211-195b13.1*	-2.41	No expression pattern	
Dr.9611.1.A1_at	*si:ch211-241e15.2*	-2.38	No expression pattern	Yes
Dr.723.1.A1_at	*cno*	-2.37	Yes, not spatially restricted	
Dr.8680.1.S1_at	*claudin j*	-2.35	Yes, OV	Yes, OV
Dr.14326.1.A1_at		-2.34		
Dr.4751.1.S1_a_at	*aldh2b*	-2.34	Yes, NC	Yes, NC
Dr.10941.1.A1_at		-2.25		
Dr.22990.1.A1_at	*wu:fj82h06*	-2.23	No expression pattern	Yes
Dr.17217.1.S1_at	*rab32*	-2.22	Yes, NC	
Dr.5325.1.A1_at	*zgc:65870*	-2.22	Yes	Yes
Dr.12707.1.A1_at	*zgc:112072*	-2.20	Yes, NC	Yes
Dr.10624.1.S1_at	*zgc:110343*	-2.18	Yes, NC	Yes, NC
Dr.23536.1.A1_at	*si:dkeyp-38g6.2*	-2.18	No expression	

			pattern	
Dr.6806.1.S1_at	*keap1a*	-2.16	Yes	Yes
Dr.22923.1.A1_at	*zgc:110355*	-2.16	Yes, NC	
Dr.5513.1.S1_at	*slc3a2*	-2.13	Yes, DNT, possible NC	
Dr.17743.1.A1_at	*zgc:66482*	-2.13	Yes, OV	Yes
Dr.12551.1.S1_at		-2.13		
Dr.20434.1.A1_at		-2.12		
Dr.3773.1.S1_at	*uap1*	-2.11	Yes, OV	
Dr.14668.1.S1_at	*gch2*	-2.11	Yes, NC	
Dr.11767.1.A1_at	*epo*	-2.10	Yes	
Dr.25465.1.A1_at	*si:dkey-15j16.2*	-2.09	Yes, not spatially restricted	
Dr.10376.1.S1_at		-2.06		
Dr.13952.1.A1_at	*zgc:63553*	-2.04	No expression pattern	
Dr.19899.1.A1_at		-2.04		
Dr.13145.1.A1_at		-2.04		
Dr.12107.1.A1_at	*ndrg1*	-2.03	Yes, OV	Yes
Dr.23838.1.S1_at	*wu:fc38f09*	-2.00	No expression pattern	

Table 55: Table of the down-regulated genes ranked by fold change identified from the combined data set.

The Affymetrix probe set identifier was used to obtain the corresponding gene symbol from the NetAffx website. Gene symbols were used to interrogate the ZFIN website for published expression patterns. Yes = a pattern was available, when this identified expression in the neural crest (NC), otic vesicle (OV) or dorsal neural tube (DNT) this has been noted, when no NC, OV or DNT was observed no sites of expression have been noted. No expression pattern = gene was present on ZFIN but no pattern was available. Yes, not spatially restricted = the expression pattern was unclear. Blank cells indicate that the gene was not present on the ZFIN website. The ISH screen column indicates if a gene was examined during the *in situ* hybridization screen. If expression was observed in the neural crest or otic vesicle this has been noted; Yes = a pattern was available, when this identified expression in the NC, OV or DNT this has been noted, when no NC, OV or DNT expression was observed, no sites of expression have been noted. Blank cells indicate the gene was not examined by *in situ* hybridization. Genes have been ranked by fold change.

Affymetrix ID	Gene Symbol	Fold Change	ZFIN Expression Pattern	ISH Screen
Dr.2596.2.A1_a_at	*mpv17*	3.14	Yes, NC	
Dr.23341.1.A1_at		2.73		
Dr.20391.1.A1_at	*LOC563577*	2.60		
Dr.17625.1.A1_at	*si:rp71-69p9.2*	2.58	No expression pattern	
Dr.10405.1.S1_at	*ca2*	2.45	Yes	
Dr.11260.1.A1_at		2.35		
Dr.15195.1.A1_at		2.32		
Dr.10488.1.A1_at	*ngfr*	2.28	No expression pattern	
Dr.12717.1.S1_at	*sulf2*	2.27	Yes	
Dr.185.1.A1_at	*wu:fa04h02*	2.22	No expression pattern	
Dr.13165.1.A1_at	*npr3*	2.19	Yes	Yes
Dr.8061.1.S1_at	*sema3d*	2.16	Yes, NC	
Dr.7441.1.A1_at	*wu:fi31c04*	2.16	No expression pattern	
Dr.8617.1.A1_at	*id:ibd5023*	2.14	Yes	
Dr.2967.1.A1_at	*hoxc8a*	2.10	Yes	Yes
Dr.19844.1.A1_at	*CH211-173B8.2*	2.09	No expression pattern	
Dr.6603.1.A1_at	*ehd2*	2.08	Yes	
Dr.5129.1.S1_at	*sesn3*	2.08	Yes	
Dr.26411.2.S1_s_at	*LOC792267* *Tnni2* *tnni2a.2* *tnni2a.4*	2.07	Yes	
Dr.13868.1.S1_at	*syt4*	2.04	Yes, possible DNT	
Dr.14522.1.S1_at	*pvalb7*	2.04	Yes	Yes
Dr.12963.1.A1_at		2.00		
Dr.18513.1.A1_at	*sccpdhb*	2.00	Yes	Yes

Table 56: Table of the up-regulated genes in the combined data set, ranked by fold change.
The Affymetrix probe set identifier was used to identify the gene symbol using the NetAffx website. Gene symbols were used to interrogate the ZFIN website for published expression patterns at 24 hpf. Table is presented in the same format as Table 55.

5.2.2.2 Identifying candidates differentially regulated in both 7.2sox10:GFP and 4.9sox10:GFP data sets.

All Affymetrix probe sets identified as significantly down-regulated in *sox10* mutants using the *t*-Test ($p<0.05$) from analysis of the two initial data sets were independently collected. Thus significantly differentially-regulated genes identified from both the *7.2sox10*:GFP and *4.9sox10*:GFP lines were collected. In this section, the differentially regulated genes were compared to identify common down-regulated or up-regulated probe sets using the intersect function on the NetAffx website.

89 down-regulated probe sets were identified as common between the two gene lists (**Table 57**). 30 of these probe sets were evaluated during the *in situ* hybridization screen, 19 of which were validated as NC or otic vesicle expressed and 11 probe sets (37 %) were false positives. This low rate of false positives suggests that the list of genes

presented in **Table 57** will provide a reliable set of genes to screen for N or otic vesicle expression. To examine this further, the list of probe sets was interrogated for expression patterns on ZFIN. This identified that 23 probe sets had a NC expression pattern and 4 had an otic vesicle expression pattern. Three probe sets displayed an expression pattern on ZFIN that suggested NCCs were marked. 39 probe sets did not have a published expression pattern, 8 had no clear expression pattern and 12 had a clear expression pattern that was not of interest. Integrating the *in situ* hybridization data from Chapter 4 and the published expression pattern data identified that 31 probe sets had a published NC or otic vesicle expression pattern. Therefore a large number of genes expressed in the cell types of interest were represented by the probe sets identified as down-regulated in both the *7.2sox10*:GFP and *4.9sox10*:GFP microarray data sets (**Table 57**). 4 genes expressed in the otic vesicle were present in **Table 57**, 3 of which were not assessed during the *in situ* hybridization screen. Thus otic vesicle expressed genes were identified by this method. Three genes had an expression pattern on ZFIN that suggested NCCs were marked but were not annotated as such, but this requires further investigation. This demonstrates the danger of assuming that the annotation of published gene expression data is comprehensive. 15 genes had a clear expression pattern determined from ZFIN or were assessed during the *in situ* hybridization screen and displayed no NC or otic vesicle expression. Thus the remaining 42 genes had no clear published expression pattern and could be investigated. The presence of a large number of genes with an expression pattern of interest contrasts with only a few genes that were false positives. This suggests the remaining candidates should be a priority for validation.

The same analysis was performed with genes that were up-regulated in both of the independent microarray data sets (**Table 58**). This identified 51 probe sets of which 3 had a clear published NC expression pattern and 3 were expressed in the otic vesicle on ZFIN. None of these 6 genes were assessed during the *in situ* hybridization screen; all therefore should be examined to identify if these genes are up-regulated in *sox10* mutant embryos. 20 genes had a clear published expression pattern but displayed no expression in the NC or otic vesicle. This suggested that the percentage of positive hits in the up-regulated data set was low, as noted in Chapter 4. In addition, 9 genes that were assessed during the *in situ* hybridization screen appeared in **Table 58**. None of these were expressed in the NC or otic vesicle thus all are false positives. This suggested that a large number of false positive genes were present in the gene list. However, 25 genes had no clear published expression pattern and could be investigated despite the likelihood that many are false positives.

Probe Set ID	Gene Symbol	ZFIN Expression Pattern	ISH Screen
Dr.4751.1.S1_a_at	*aldh2b*	Yes, NC	Yes, NC
Dr.4751.2.A1_at	*aldh2b*	Yes, NC	Yes, NC
Dr.25417.1.A1_at	*ap3m2*	No expression pattern	
Dr.2855.1.A1_a_at	*atp6v1a*	Yes	Yes, NC
Dr.186.1.S1_at	*atp6v1ba*	Yes, possible NC expression	
Dr.14783.1.S1_at	*atp6v1c1*	Yes, not spatially restricted	
Dr.7966.1.S1_at	*atp6v1e1*	No expression pattern	Yes, NC
Dr.370.1.S1_at	*atp6v1g1*	Yes, possible NC expression	
Dr.25257.1.A1_at	*bhlhb3l*	No expression pattern	Yes
Dr.7400.1.A1_at	*csnk1e*	Yes, OV	
Dr.5040.1.S1_at	*cyb5a*	Yes, NC	
Dr.10336.1.S1_at	*dct*	Yes, NC	Yes, NC
Dr.6362.1.A1_at	*depdc6*	No expression pattern	
Dr.20850.1.S1_at	*fabp7a*	Yes	Yes
Dr.4801.1.S1_at	*gart*	Yes, NC	
Dr.17596.1.A1_at	*LOC100000591*		
Dr.10530.1.S1_at	*LOC100002236*		
Dr.9225.1.A1_at	*LOC100148138*		
Dr.2072.1.A1_at	*LOC553451*		
Dr.14390.1.S1_at	*LOC557898*		
Dr.752.1.A1_at	*LOC562425*		
Dr.12107.1.A1_at	*ndrg1*	Yes, OV	Yes
Dr.19416.1.S1_at	*otomp*	Yes, OV	Yes, OV
Dr.4171.1.A1_at	*pah*	Yes, NC	Yes, NC
Dr.983.1.S1_at	*paics*	Yes, NC	Yes, NC
Dr.360.1.A1_at	*pla2g15*	No expression pattern	
Dr.26358.1.A1_at	*ppat*	Yes, NC	
Dr.8045.1.S1_at	*prtfdc1*	Yes, not spatially restricted	
Dr.14001.1.A1_at	*pttg1ip*	Yes, NC	
Dr.14001.2.S1_at	*pttg1ip*	Yes, NC	
Dr.17217.1.S1_at	*rab32*	Yes, NC	
Dr.10292.1.S1_at	*rbp4l*	Yes, NC	Yes, NC
Dr.518.1.A1_at	*sb:cb319*	Yes	Yes
Dr.12651.1.S1_at	*sh3glb2*	Yes	
Dr.1823.2.S1_at	*si:ch211-150c22.2*	Yes, not spatially restricted	
Dr.6142.1.A1_at	*si:dkey-180p18.9*	Yes, not spatially restricted	

Dr.25685.1.A1_at	*si:dkey-31f5.7*	Yes, NC	
Dr.21747.1.A1_at	*si:dkey-94n12.4*	No expression pattern	
Dr.14747.1.A1_at	*slc2a15b*	Yes	Yes, NC
Dr.20896.1.S1_at	*slc2a15b*	Yes	Yes, NC
Dr.4885.1.S1_at	*slc34a2a*	Yes, not spatially restricted	
Dr.24728.1.S1_at	*slc37a2*	No expression pattern	
Dr.5513.1.S1_at	*slc3a2*	Yes, possible NC, DNT	
Dr.13477.1.A1_at	*slc9a8*	Yes, not spatially restricted	
Dr.20.1.S1_at	*snai2*	Yes, NC	
Dr.15842.1.A1_at	*syngr1*	Yes, NC	
Dr.3773.1.S1_at	*uap1*	Yes, OV	
Dr.14652.1.A1_at	*ube2k*	Yes	
Dr.18689.1.A1_at	*vac14*	Yes, not spatially restricted	
Dr.18964.1.A1_at	*vac14*	Yes, not spatially restricted	
Dr.4248.1.A1_at	*wu:fa98e08*	No expression pattern	Yes
Dr.4612.1.A1_at	*wu:fc31e04*	No expression pattern	Yes, NC
Dr.23838.1.S1_at	*wu:fc38f09*	No expression pattern	
Dr.5170.2.A1_at	*wu:fc41e04*	No expression pattern	
Dr.3933.1.A1_at	*wu:fc46b01*	No expression pattern	Yes
Dr.9746.7.A1_at	*wu:fc51f04*	No expression pattern	
Dr.2268.1.A1_at	*wu:fc64c05*	No expression pattern	
Dr.22266.1.A1_at	*wu:fe48e11*	No expression pattern	
Dr.8251.1.A1_at	*wu:fi05a09*	Yes, NC	Yes, NC
Dr.11035.1.A1_at	*wu:fj35f12*	No expression pattern	
Dr.7226.1.S1_at	*wu:fj65h10*	No expression pattern	
Dr.14073.1.A1_at	*wu:fj78f01*	Yes	
Dr.8594.1.S1_at	*zgc:100919*	Yes, NC	Yes, NC
Dr.4057.1.S1_at	*zgc:101045*	No expression pattern	
Dr.10145.1.A1_at	*zgc:110003*	No expression pattern	
Dr.10624.1.S1_at	*zgc:110343*	Yes, NC	Yes, NC
Dr.10624.2.S1_at	*zgc:110343*	Yes, NC	Yes, NC
Dr.10624.2.S1_a_at	*zgc:110343*	Yes, NC	Yes, NC
Dr.12707.1.A1_at	*zgc:112072*	Yes, NC	Yes
Dr.13330.1.A1_at	*zgc:112153*	No expression pattern	
Dr.9036.1.A1_at	*zgc:136982*	No expression pattern	
Dr.4838.1.A1_at	*zgc:158673*	Yes	Yes
Dr.4868.1.A1_at	*zgc:174888*	Yes	

Dr.5325.1.A1_at	*zgc:65870*	Yes	Yes
Dr.9944.1.A1_at	*zgc:91968*	Yes, NC	
Dr.14275.1.A1_at	*zgc:92185*	No expression pattern	
Dr.10941.1.A1_at			
Dr.13919.1.S1_at			No
Dr.14838.1.A1_at			
Dr.16789.1.A1_at			No
Dr.17577.1.A1_at			
Dr.21407.1.A1_at			No
Dr.26012.1.A1_at			
Dr.26101.1.A1_at			No
Dr.26254.1.A1_at			
Dr.26522.1.A1_at			
Dr.3972.1.S1_at			Yes, NC
Dr.4836.1.A1_at			
Dr.8674.1.A1_at			

Table 57: Table of the probe sets identified as down-regulated in both the *7.2sox10*:GFP and *4.9sox10*:GFP data sets.
Probe set Affymetrix IDs were annotated using NetAffx. Table is presented as per Table 55 except genes are ordered alphabetically**.**

Probe Set ID	Gene Symbol	ZFIN Expression Pattern	ISH Screen
Dr.10405.1.S1_at	*ca2*	Yes	
Dr.25673.2.S1_at	*cdc25*	No expression pattern	
Dr.14422.3.A1_x_at	*cdon*	Yes, OV	
Dr.22283.1.S1_at	*cnn3a*	Yes, NC	
Dr.15055.1.S1_at	*cxcr4a*	Yes	
Dr.574.2.S1_at	*dlb*	Yes	
Dr.21809.1.S1_at	*fbxl14*	Yes, not spatially restricted	
Dr.21414.1.A1_at	*hoxb9a*	Yes, not spatially restricted	
Dr.2967.1.A1_at	*hoxc8a*	Yes	Yes
Dr.8617.1.A1_at	*id:ibd5023*	Yes	
Dr.17792.2.A1_at	*im:7163473*	Yes, not spatially restricted	
Dr.23153.1.S1_at	*LOC100000500*		Yes
Dr.11563.1.A1_at	*LOC553231*		
Dr.13854.1.S1_at	*LOC569855*		
Dr.14994.1.A1_at	*mapk14b*	Yes, not spatially restricted	
Dr.24241.1.S1_at	*mdkb*	Yes, OV	
Dr.1680.1.S1_at	*meis3*	Yes, NC	

Dr.4771.1.S1_at	*ndufb9*	Yes, not spatially restricted	
Dr.10488.1.A1_at	*ngfr*	No expression pattern	
Dr.13165.1.A1_at	*npr3*	No	Yes
Dr.14203.1.S1_at	*ppp1r14a*	Yes	
Dr.14522.1.S1_at	*pvalb7*	Yes	Yes
Dr.24815.1.S1_at	*rpl13*	Yes, not spatially restricted	
Dr.1691.14.S1_at	*sb:cb8*	Yes	
Dr.9438.1.A1_at	*si:dkey-218l8.1*	No expression pattern	
Dr.7094.1.A1_at	*si:dkey-24l11.4*	Yes, not spatially restricted	
Dr.4624.1.A1_at	*si:dkey-32e6.1*	Yes, not spatially restricted	
Dr.6063.1.A1_at	*si:dkey-73n10.1*	No expression pattern	Yes
Dr.25679.1.S1_at	*slc1a3*	Yes, OV	
Dr.7855.2.S1_at	*sncb*	Yes	
Dr.12717.1.S1_at	*sulf2*	Yes	
Dr.13868.1.S1_at	*syt4*	Yes, possible DNT	
Dr.25500.1.S1_at	*vamp1*	Yes, possible DNT	Yes
Dr.26450.1.S1_at	*vegfc*	No	
Dr.18141.2.S1_at	*wu:fa99c11*	No expression pattern	
Dr.21543.1.A1_at	*wu:fc41f11*	No expression pattern	
Dr.22087.1.A1_at	*wu:fd12f01*	No expression pattern	Yes
Dr.1094.1.A1_at	*wu:fk49e12*	No expression pattern	Yes
Dr.15362.1.S1_at	*zgc:110256*	Yes, not spatially restricted	
Dr.25338.1.A1_at	*zgc:123165*	Yes	
Dr.17582.1.A1_at	*zgc:152916*	No expression pattern	
Dr.3936.1.A1_at	*zgc:162730*	No expression pattern	Yes
Dr.23549.1.A1_at	*zgc:92231*	Yes, OV	
Dr.13091.1.S1_at	*zgc:92878*	Yes, not spatially restricted	
Dr.10689.1.S1_at	*zic2b*	Yes, possible DNT	
Dr.11232.1.A1_at			
Dr.12819.1.A1_at			
Dr.14210.1.A1_at			
Dr.14213.1.A1_at			
Dr.15195.1.A1_at			
Dr.2693.1.A1_at			

Table 58: Table of the probe sets identified as up-regulated in both the *7.2sox10*:GFP and *4.9sox10*:GFP data sets.
Probe set Affymetrix IDs were annotated using NetAffx. Table is presented as per Table 55 except that genes are ordered alphabetically.

5.2.2.3 Examining candidates significantly differentially regulated after methods to control for multiple testing were applied.

A small number of genes were identified as significantly differentially regulated in the combined data set after controlling for multiple testing using the FDR method. In a list of

genes identified as differentially regulated under FDR criteria, it would be expected that a known number of genes are false positives (Cui and Churchill, 2003). 32 down-regulated genes were significant using the FDR method (p<0.05) (**Table 59**). 18 of these genes had a clear published NC or otic vesicle expression pattern on ZFIN or were validated during the *in situ* hybridization screen. 26 of the 32 genes were examined during the *in situ* hybridization screen, identifying 15 NC or otic vesicle expressed genes and 11 false positives (11/26 = 42 %). The false positive percentage is considerably higher than the expected 5 % rate (at p<0.05 significance level). This left 6 genes that had not been examined during the *in situ* hybridization screen and did not have a clear expression pattern on ZFIN. These genes present attractive targets for validation although it should be noted that *s100b* present in this list of genes could not be amplified by PCR from 24 hpf cDNA.

For the up-regulated genes significant after applying the FDR method (p<0.05), 7 genes were identified. Of these, 6 had a clear expression pattern on ZFIN, 3 of which were also examined during the *in situ* hybridization screen. None of these genes had a NC or otic vesicle expression pattern. Thus all genes examined by *in situ* hybridization were false positives. One gene had no clear expression pattern on ZFIN but was examined during the *in situ* hybridization screen but was not validated. Thus this list of up-regulated target genes contains no obvious candidates to prioritise for validation.

Affymetrix ID	Gene Symbol	FDR	ZFIN Expression Pattern	ISH Screen
Dr.3933.1.A1_at	*wu:fc46b01*	0.0028	No expression pattern	Yes
Dr.25257.1.A1_at	*bhlhb3l*	0.0028	No expression pattern	Yes
Dr.2855.1.A1_a_at	*atp6v1a*	0.0034	No expression pattern	Yes, NC
Dr.10292.1.S1_at	*rpb4l*	0.0036	Yes, NC	Yes, NC
Dr.3972.1.S1_at		0.0036		Yes, NC
Dr.8594.1.S1_at	*zgc:100919*	0.0051	Yes, NC	Yes, NC
Dr.13919.1.S1_at		0.0100		Yes
Dr.4612.1.A1_at	*wu:fc31e04*	0.0111	No expression pattern	Yes, NC
Dr.518.1.A1_at	*sb:cb319*	0.0122	Yes	Yes
Dr.18194.1.A1_at	*pik3c3*	0.0125	Yes, not spatially restricted	Yes
Dr.7966.1.S1_at	*atp6v1e1*	0.0125	No expression pattern	Yes, NC
Dr.12107.1.A1_at	*ndrg1*	0.0153	Yes, OV	Yes
Dr.4171.1.A1_at	*pah*	0.0174	Yes, NC	Yes, NC
Dr.4751.1.S1_a_at	*aldh2b*	0.0174	Yes, NC	Yes, NC
Dr.6142.1.A1_at	*si:dkey-180p18.9*	0.0174	Yes, not spatially restricted	
Dr.14747.1.A1_at	*slc2a15b*	0.0174	Yes, not spatially restricted	Yes, NC
Dr.10624.2.S1_a_at	*zgc:110343*	0.0244	Yes, NC	Yes, NC
Dr.983.1.S1_at	*paics*	0.0244	Yes, NC	Yes, NC
Dr.8251.1.A1_at	*wu:fi05a09*	0.0244	Yes, NC	Yes
Dr.17577.1.A1_at		0.0244		
Dr.370.1.S1_at	*atp6v1g1*	0.0251	Yes	
Dr.481.1.S1_at	*atic*	0.0251	Yes, NC	Yes, NC
Dr.16789.1.A1_at		0.0251		Yes
Dr.20896.1.S1_at	*slc2a15b*	0.0262	Yes, not spatially restricted	Yes, NC
Dr.10336.1.S1_at	*dct*	0.0281	Yes, NC	Yes, NC
Dr.25465.1.A1_at	*si:dkey-15j16.2*	0.0301	Yes, not spatially restricted	
Dr.17743.1.A1_at	*zgc:66482*	0.0301	Yes, OV	Yes
Dr.2729.1.A1_at	*wu:fc54b10*	0.0344	No expression pattern	Yes
Dr.14787.1.A1_at	*s100b*	0.0344		
Dr.4751.2.A1_at	*aldh2b*	0.0410	Yes, NC	Yes, NC
Dr.26522.1.A1_at		0.0410		
Dr.26101.1.A1_at		0.0434		Yes

Table 59: Table of probe sets significantly down-regulated after controlling for multiple testing.

The Affymetrix probe set identifiers significant by FDR were annotated using the NetAffx website. Gene symbols were used to interrogate the ZFIN website for published expression patterns. Data is presented as per Table 55 except genes are ranked by FDR adjusted *p*-value.

Affymetrix ID	Gene Symbol	FDR	ZFIN Expression Pattern	ISH Screen
Dr.13165.1.A1_at	*npr3*	0.0034	Yes	Yes
Dr.6063.1.A1_at	*si:dkey-73n10.1*	0.0054	No expression pattern	Yes
Dr.2967.1.A1_at	*hoxc8a*	0.0088	Yes	Yes
Dr.10405.1.S1_at	*ca2*	0.0174	Yes	
Dr.25673.2.S1_at	*cdc25*	0.0251	Yes	
Dr.14522.1.S1_at	*pvalb7*	0.0251	Yes	Yes
Dr.26450.1.S1_at	*vegfc*	0.0410	Yes	

Table 60: Table of probe sets significantly up-regulated after controlling for multiple testing.

The Affymetrix probe set identifiers significant by FDR were annotated using the NetAffx website. Gene symbols were used to interrogate the ZFIN website for published expression patterns. Data is presented as per Table 55 except genes are ranked by FDR adjusted *p*-value.

5.2.2.4 Comparing the methods used to identify high priority genes for validation.

The results from the three methods utilised in this section to generate lists of genes are summarised in **Table 61**. This table includes the same results for the blind *in situ* screen performed on the *7.2sox10*GFP *t*-Test data for comparison. In addition, the same data was gathered for the 7.2+4.9 combined data set. From these methods, the number of probe sets that were assessed during the *in situ* hybridization screen as true positives (NC or otic vesicle expressed and differentially regulated in *sox10* mutant embryos) or false positives are given. This provides an indication of how an *in situ* hybridization screen might perform on the whole gene list. This data has also been combined with that of published expression patterns on the ZFIN website to indicate the total number of likely positive results that are present in a gene list. This also enabled the number of genes that lack any published expression data to be determined as these candidates would be targeted to yield novel NC and otic vesicle expressed genes when examined by *in situ* hybridization.

For the down-regulated gene lists, all three methods appeared to identify a higher proportion of positive hits than were identified during the blind *in situ* hybridization screen and a smaller proportion of false positives were also identified (**Table 61**). Thus further investigation of the genes identified by these methods may be more valuable than continuing a blind screen on the *7.2sox10*:GFP data set. Alternatively performing a blind screen on the combined data set (focusing on the most down-regulated genes not investigated in Chapter 4) is likely to reavel more genes of interest. For two of these analyses, fold change and FDR, the combined data set was examined rather than the *7.2sox10*:GFP which was the basis for the blind *in situ* hybridization screen. This may account for some of the differences detected in positive and false positive hit rates. A number of poorly characterised probe sets (only annotated with an Affymetrix identifier) have not been investigated in the data sets summarised in **Table 61**. When these probe sets were examined by *in situ* hybridization they typically had no clear or specific gene expression pattern (false positive) thus the hit rates predicted in **Table 61** are likely to be slightly high. However it should also be noted that some poorly characterised probe sets, such as Dr.3972.1.S1_at, were expressed in the NC and validated as down-regulated in *sox10* mutant embryos. Based on the results presented in **Table 61**, the method that produced the best true positive hit rate and the lowest false positive hit rate was the comparison of down-regulated genes in the two independent microarray data sets. This appears to be the gene list that should be prioritised for validation. For many studies a comparison of two different but similar sets of data will not be possible. Controlling for multiple testing (by the FDR method) also generated a list of genes that contained many true positives and a reduced number of false positives compared to the screen performed in Chapter 4. Thus this method could also be used to select genes for validation from

microarray data. While cross-referencing the *in situ* hybridization screen data with expression patterns on the ZFIN website it was noticed that 3 genes had an expression pattern of interest on ZFIN but were not validated during the screen. The published otic vesicle expression of *ndrg1* and *zgc:66482* and the NC expression pattern of *zgc:112072* were not detected. These genes should be re-evaluated to confirm the *in situ* hybridization screen result.

For the up-regulated gene lists, neither of the two otic vesicle expressed genes validated during the *in situ* hybridization screen were identified by the three methods. The lack of known up-regulated markers on the Affymetrix microarray chip makes analysis of the information presented in this chapter difficult and no gene list can be prioritised for validation. However, several genes known to be NC or otic vesicle were highlighted during these analyses. To increase the range of known up-regulated marker genes available, these known genes should be prioritised for examination by *in situ* hybridization. This could facilitate the interpretation of the up-regulated data sets which has proved problematic at all stages of this project thus far.

During the *in situ* hybridization screen only five ear expressed genes were identified and validated as differentially regulated from the genes examined. Very few known otic vesicle expressed genes were identified by the various methods utilised here to sort microarray data. This was despite a ZFIN anatomy search for the term “otic vesicle” returning 520 expressed genes at the same time point that 130 genes are NC expressed. Thus the screen has, up to this point, not proved as successful at identifying otic vesicle expressed genes when compared to NC expressed genes.

	Method	Number of ISH positives	Number of ISH false positives	Number of known NC or otic vesicle expressed genes	Number of genes lacking a clear expression pattern
Down-regulated genes	7.2 data set *t*-test	28/92 (30.4 %)	64/92 (69.6 %)	31/92 (33.7 %)	-
	Fold Change	13/30 (43.3 %)	17/30 (56.7 %)	20/37 (54.1 %)	18
	Identified in two independent data sets	19/30 (63.3 %)	11/30 (36. 7 %)	31/46 (67.4 %)	40 (+ 3 possible NC expressed genes)
	FDR	15/26 (57.7 %)	11/26 (42.3 %)	18/27 (66.7 %)	5
	7.2+4.9 Combined data set	17/30 (56.7 %)	13/30 (43.3 %)	33/100 (33.3 %)	47
Up-regulated genes	7.2 data set *t*-test	2/47 (4.3 %)	45/47 (95.7 %)	2/47 (4.3 %)	-
	Fold Change	0/4 (0 %)	4/4 (100 %)	2/13 (15.4 %)	10
	Identified in two independent data sets	0/9 (0 %)	9/9 (100 %)	6/26 (23.1 %)	25
	FDR	0/4 (0 %)	4/4 (100 %)	0/7 (0 %)	0
	7.2+4.9 Combined data set	0/9 (0 %)	9/9 (100 %)	6/50 (12 %)	34

Table 61: Summary of the methods employed to identify genes for validation.
Results from the three methods described in this section and the *t*-Test used to select genes for *in situ* hybridization validation in Chapter 4 are shown. From the probe sets identified by each method, the numbers of probe sets assessed by in situ hybridization in Chapter 4 are split into positive and negative (false positive) results. This data was combined with published ZFIN data to identify the number of probe sets identified by each method that are NC or otic vesicle expressed. The number of probe sets that were not assessed during the *in situ* hybridization screen and do not have a clear published expression pattern are also given.

5.2.3 Identifying interesting genes from microarray data by function.

5.2.3.1 Identifying Genes of interest by GO Term using Babelomics.

To highlight genes with a function of interest identified as differentially regulated in the combined 7.2+4.9 microarray data set, the Babelomics FatiGO tool was used. A list of all significantly (*t-Test* p-*value* <0.05) down or up-regulated genes was loaded into the tool for analysis of GO biological process and molecular function terms. The genes represented by a GO term of interest were then evaluated.

5.2.3.1.1 Transcription Factors

A key aim of this microarray project was to identify Sox10 dependent transcription factors that play important roles in NC development. All of the genes identified as down-regulated in the combined data set were sorted by GO term using Babelomics. Of the 640 down-regulated genes, 190 were annotated with a molecular function GO term. Two level 5 GO terms appeared to be likely candidates to facilitate identifying transcription factor genes. 7 genes were annotated by the GO term "Transcription factor activity" (GO:0003700), *esr1, foxd3, mtx1, poll, rorab, runx2b* and *wu:fa05a04* (**Table 62**). All of these genes except *runx2b* were also annotated by the level 5 GO term "Sequence specific DNA binding" (GO:0043565). A parent (level 4) GO term) of both of these terms, "DNA binding" (GO:0003677), was associated with 19 genes including all the genes identified as potential transcription factor genes so far (**Table 62**). A number of genes at this annotation level were not transcription factors, for example, *paics* is an enzyme involved in purine biosynthesis (Li et al., 2007). Thus the level 4 GO term covers a very broad range of genes, not simply transcription factors. The level 5 GO term "Transcription factor activity" identified a number of transcription factor genes including *foxd3*. The majority of these genes had a published expression pattern and did not display NC or otic vesicle expression. However, the gene *wu:fa05a04* had no expression pattern and is therefore a candidate for validation. By adopting a more specific GO term, some genes which were not transcription factors were excluded but some genes such as the two basic helix-loop-helix domain containing genes (*bhlhb2* and *bhlhb3l*) were excluded. The helix-loop-helix domain is a DNA binding domain found in proteins that function as transcription factors, including the melanocyte transcription factor *Mitf* (Hodgkinson et al., 1993). Thus *bhlhb2* and *bhlhb3l* are likely to function as transcription factors. Depending on which level of GO term was selected for analysis, some non-transcription factor genes were selected or with a more specific GO term some likely transcription factor genes were excluded.

The key melanocyte transcription factor *mitfa* was surprisingly not annotated by the GO Terms examined so far. Thus to facilitate identifying potential transcription factor genes, the GO terms associated with *mitfa* were identified using Babelomics. The GO terms that appeared to be linked to the function of *mitfa* as a transcription factor were as follows: GO Molecular function level 5 "Transcriptional activator activity" (GO:0016563), GO Biological process level 5 "Transcription" (GO:0006350), GO biological process level 7 "Transcription, DNA-dependent" (GO:0006351) and "Regulation of transcription" (GO:0045449). The combined down-regulated Babelomics data was examined for these GO terms (**Table 63**). Along with the key melanocyte transcription factor *mitfa*, the NC transcription factor *foxd3* was also annotated with all of these GO terms. The other gene annotated with all of these GO terms was *wu:fa05a04*, which was also identified previously (**Table 62** and **Table 63**). This again suggests that this gene represents an

interesting candidate for validation. As with the GO term analysis presented in **Table 62**, some genes annotated with a GO term of interest are not transcription factors. For example, from the 20 genes presented in **Table 63**, *polr3f* and *polr2c* both encode subunits of RNA polymerase. Manual selection of transcription factors from these generated gene lists would therefore be required to address the specific aim of identifying transcription factors for validation.

Affymetrix ID	Gene Symbol	DNA Binding	Transcription Factor Activity	Sequence Specific DNA Binding	NC or otic vesicle expressed	ISH Result
Dr.10717.1.S1_at	*esr1*	✓	✓	✓	No	
Dr.17137.1.S1_at	*rorab*	✓	✓	✓	No	
Dr.20.1.S1_at	*snai2*	✓			Yes, NC	
Dr.21045.1.S1_at	*rabgef1*	✓			Not spatially restricted	
Dr.24983.1.S1_at	*polr3f*	✓			Not spatially restricted	
Dr.25257.1.A1_at	*bhlhb3l*	✓				No
Dr.2654.3.S1_a_at	*rfx2*	✓			No	
Dr.3180.1.A1_at	*bhlhb2*	✓			Yes, early NC	No
Dr.462.1.A1_at	*wu:fa05a04*	✓	✓	✓		
Dr.5094.2.A1_at		✓				
Dr.5418.2.S1_at	*top2a*	✓				
Dr.5740.1.S1_at	*neurod2*	✓				
Dr.590.1.S1_at	*foxd3*	✓	✓	✓	Yes, NC	Yes, NC
Dr.6031.2.A1_at	*poll*	✓	✓	✓	Not spatially restricted	
Dr.7734.1.S1_a_at	*mcm3l*	✓			Not spatially restricted	
Dr.8324.1.S1_at	*mxtx1*	✓	✓	✓	No	
Dr.9758.1.S1_at	*polr2c*	✓			Not spatially restricted	
Dr.983.1.S1_at	*paics*	✓			Yes, NC	Yes, NC
DrAffx.2.6.S1_s_at	*runx2b*	✓	✓		No	

Table 62: Summary of potential transcription factor genes as identified by GO Term. All genes identified with appropriate GO Terms relating to possible transcription factor activity are listed. Where genes have a published expression pattern on ZFIN this has been noted, blank cells indicate that no pattern was available. If a gene was examined during the *in situ* hybridization screen this has been noted (No = not validated, Yes = validated). NC = neural crest expression was observed.

Probe Set ID	Gene Symbol	Tran-scription	Regulation of Tran-scription	Tran-scription, DNA-dependent	Tran-scriptional Activator	NC or otic vesicle expressed	ISH Result
Dr.10717.1.S1_at	*esr1*	✓	✓	✓		No	
Dr.12260.1.A1_at	*rab7*	✓	✓	✓		Yes, NC	
Dr.14855.1.S1_a_at	*rab35*	✓	✓	✓			
Dr.17137.1.S1_at	*rorab*	✓	✓	✓		No	
Dr.20.1.S1_at	*snai2*	✓	✓	✓		Yes, NC	
Dr.24248.1.S1_at	*jmjd2al*	✓	✓	✓			
Dr.24983.1.S1_at	*polr3f*	✓	✓	✓		Not spatially restricted	
Dr.25257.1.A1_at	*bhlhb3l*	✓	✓	✓			No
Dr.2654.3.S1_a_at	*rfx2*	✓	✓	✓		No	
Dr.3148.2.A1_at		✓	✓				
Dr.3180.1.A1_at	*bhlhb2*	✓	✓	✓		Yes, early NC	No
Dr.462.1.A1_at	*wu:fa05a04*	✓	✓	✓	✓		
Dr.4874.1.S1_at	*rab1a*	✓	✓	✓		Not spatially restricted	
Dr.5094.2.A1_at		✓	✓	✓			
Dr.5740.1.S1_at	*neurod2*	✓	✓	✓			
Dr.590.1.S1_at	*foxd3*	✓	✓	✓	✓	Yes, NC	Yes, NC
Dr.8080.1.S1_at	*mitfa*	✓	✓	✓	✓	Yes, NC	Yes, NC
Dr.8324.1.S1_at	*mxtx1*	✓	✓	✓		No	
Dr.9758.1.S1_at	*polr2c*	✓	✓	✓		Not spatially restricted	
DrAffx.2.6.S1_s_at	*runx2b*	✓	✓	✓		No	

Table 63: Summary of potential transcription factor genes identified by GO Term.
All genes identified with appropriate GO Terms relating to possible transcription factor activity are listed. Where genes have a published expression pattern on ZFIN this has been noted, blank cells indicate that no pattern was available. If a gene was examined during the *in situ* hybridization screen this has been noted (No = not validated, Yes = validated). NC = neural crest expression was observed.

5.3 Discussion

5.3.1 Discussion on which data set presents the best list of genes for future validation.

The microarray analysis results for genes known to be down-regulated in the NC of *sox10* mutant embryos identified from the *in situ* hybridization screen were compared across the three microarray data sets. This showed that the *4.9sox10*:GFP data set often failed to correctly identify known down-regulated genes, this data set therefore generated a number of false negatives. As this screen has aimed to be inclusive to identify as many genes of interest as possible, false negatives are not desirable. The combined data set identified most of the known down-regulated NC expressed genes that were validated during the *in situ* hybridization screen (**Table 47**). In order to compare these data sets while not relying on genes picked from the *7.2sox10*:GFP data set, the results for a number of key NC marker genes were compared across all three data sets (**Table 48**). Again the *4.9sox10*:GFP data set performed worst, the *7.2sox10*:GFP data set identified 6/11 probe sets as down-regulated while the combined data set identified all genes as down-regulated. This suggested that the combined data identified down-regulated genes most efficiently with the lowest rate of false negative results. This conclusion is based on a limited number of NC expressed genes, not all of which are known to be down-regulated in *sox10* mutant embryos but based on the timing of expression would be expected to be Sox10 dependent. To determine comprehensively the best data set to use when selecting interesting genes for future validation the 100 most down-regulated genes in the *4.9sox10*:GFP and combined data sets would need to be examined by *in situ* hybridization. Then the performance of the three data sets could be accurately compared. This would require a significant extension of the *in situ* hybridization screening project to examine a large number of genes not yet assessed. This approach would be likely to identify a number of novel NC and otic vesicle expressed genes that were not examined during the *in situ* hybridization screen to date. Despite this, the strong performance of the combined data set in all the above tests suggests that this data set is the most reliable. The combined data set is the data set that should be used to address if specific genes, gene families or genes with specific functions of interest, for example transcription factors, are likely to be differentially regulated in *sox10* mutant embryos.

It was expected that differentially regulated NC genes would be detected in all three data sets as NCCs are labelled equally with GFP in the *7.2sox10*:GFP and 4.9*sox10*:GFP transgenic lines. The identification of otic vesicle genes was expected to be less effective in the *4.9sox10*:GFP data set because the otic epithelium is only weakly labelled with GFP in this line, and also therefore, to some extent in the combined data set, than the *7.2sox10*:GFP data set. Comparing the otic vesicle expressed genes identified during the *in situ* hybridization screen across all three microarray data sets suggested that

this was true. Only a few of the otic vesicle expressed genes validated from the *7.2sox10*:GFP data set were also identified as differentially regulated in the other two data sets. Genes known to be up-regulated in the otic vesicle of *sox10* mutant embryos were also examined (Dutton et al., 2009). All three data sets failed to identify the majority of these genes as up-regulated. This result suggested that otic vesicle expressed genes that are differentially regulated in *sox10* mutants were not readily or consistently designated as such in any of the data sets. In contrast to this, a number of genes were validated during the *in situ* hybridization screen, thus otic vesicle expressed genes could be identified from the *7.2sox10*:GFP data set. Otic vesicle expressed genes tend to only show small differences in gene expression between WT and *sox10* mutant embryos by *in situ* hybridization (Dutton et al., 2009). As a result, otic vesicle expressed genes may still be detected as differentially expressed by microarray analysis but perhaps not within the statistically most differentially regulated genes. Alongside this, the paucity of otic vesicle markers known to be differentially regulated in *sox10* mutants made it difficult to compare data sets and form a conclusion. Thus it was not clear which data set provided the best source to select otic vesicle expressed genes for validation. The original rationale for performing microarray analysis on the two transgenic lines was that genes differentially regulated in the *7.2sox10*:GFP data set would include NC and otic vesicle expressed genes while the *4.9sox10*:GFP data set would only include NC expressed genes. Thus subtracting the genes differentially regulated in the *4.9sox10*:GFP data set from the *7.2sox10*:GFP data set would reveal otic epithelium expressed genes. In reality, the high rate of false positives identified as differentially regulated in the *7.2sox10*:GFP data set and therefore probably also in the *4.9sox10*:GFP data set rendered this impossible.

5.3.2 Regarding the use of published expression data.

A large amount of gene expression data is available to zebrafish researchers on the ZFIN website from both published papers and directly submitted data, chiefly from a large scale gene expression screen by Thisse et al., 2004. Such data, when available, can indicate if a gene identified by microarray analysis is expressed in the tissue or cell type of interest. Previous work has used this resource to validate candidates identified from microarray data (Cerda et al., 2009). This cannot show if gene expression in the tissue or cell type of interest shown in a WT embryo is altered in a mutant embryo. Thus for this project published data can only indicate if a gene is likely to be of interest; *in situ* hybridization would still be required to validate a gene. However, all genes identified as NC or otic vesicle expressed during the *in situ* hybridization screen were validated as differentially expressed (Chapter 4). As no up-regulated NC expressed genes were identified, it remains to be seen if up-regulated NC expressed genes are always up-regulated in *sox10* mutant embryos. However, it was deemed very likely during this chapter that all NC and

otic vesicle expressed genes at 24 hpf identified as differentially regulated by microarray analysis were differentially regulated in *sox10* mutant embryos.

The Thisse et el. expression pattern screens form a very useful collection as they typically show images of whole embryos over a range of embryonic stages so that expression or lack of expression in cells or tissues of interest can be observed. In addition sites of expression are usually annotated for clarity. In rare circumstances, an expression pattern that appears to be consistent with NC expression has not been annotated. Thus visual examination of the patterns online is preferable to relying on annotation. Occasionally the Thisse et el. expression pattern is not clear, often annotated as not spatially restricted or a similar term. This may refer to a gene ubiquitously expressed in the whole embryo. During the *in situ* hybridization screen one such gene was examined and NC expression was identified. Thus while these genes may in fact be ubiquitously expressed, on rare occasions a specific expression pattern that differs from the published pattern can be obtained. When no Thisse et el. pattern is available, but expression patterns have been published, focus is on the tissue or cell type of interest. It is therefore difficult to determine if the gene is also expressed elsewhere in the embryo. The published data may also not cover the embryonic stage of interest. Unless published data explicitly stated that genes were expressed or not expressed in the NC or otic vesicle at approximately 24 hpf then it was assumed this had not been assessed. Thus Thisse et el. patterns were the most useful source of expression pattern data. Therefore a researcher could use published data to refine and prioritise a list of genes identified by microarray analysis. Genes with a well defined expression pattern in the tissue of interest would be selected for validation while those with a clear pattern that does not display expression in the tissue of interest would be the lowest priority for validation. Genes with no known published expression pattern would also be prioritised as this could lead to the validation of genes not previously associated with the biological process under investigation. Genes with a published but unclear expression pattern would form a group prioritised behind those genes without an expression pattern. This use of published data would enable a researcher to focus their resources in the validation of microarray data.

5.3.3 Examining the methods applied to refine the number of candidates generated from microarray data

The simplest measure used to rank microarray data is fold change (Cui and Churchill, 2003). Fold change is a measure of the change in the expression level of a gene between experimental samples. Thus in this project, genes with a large fold change should show a clear difference in expression patterns between WT and mutant embryos. To identify genes as differentially regulated or not, an arbitrary fold change cut off is selected. As no

statistical test is performed to generate fold change results, genes have no associated value that can indicate a level of confidence in the designation of differentially regulated or not differentially regulated (Cui and Churchill, 2003). For this chapter, a cut off of a twofold difference was selected without any prior examination of the data. A different cut off could be selected after examining the data for the performance of marker genes known to be differentially regulated or to include an appropriate number of genes for validation. Thus this method of ranking genes can be adapted to suit the data and the needs of the investigator. A large number of down-regulated NC and otic vesicle expressed genes were designated as differentially regulated using fold change (**Table 55** and **Table 56**). Thirty of these genes were assessed during the *in situ* hybridization screen and 13 were validated. This represented a higher hit rate than was achieved during the blind *in situ* hybridization screen in Chapter 4. However the entire list of genes identified by fold change would need to be examined to verify this. Only one of the validated differentially regulated otic vesicle expressed genes, *cldnj*, was identified by fold change thus the majority of genes with the largest fold changes were NC expressed. This suggested that this method may preferentially identify NC expressed genes. Otic vesicle expressed genes tend to only show small differences in gene expression between WT and *sox10* mutant embryos by *in situ* hybridization (Dutton et al., 2009). Therefore these genes would have small fold changes. Although it should be noted that all analysis methods identified a greater number of NC expressed genes than otic vesicle expressed genes. There were still a large number of false positive genes identified after ranking by fold change. That so many genes (30/55) identified by fold change as down-regulated were assessed in the 100 most down-regulated genes during the validation screen suggest that similar genes were identified by both methods. In contrast to the down-regulated genes, very few of the up-regulated genes identified by fold change were examined during the *in situ* hybridization screen. This suggests that very different results were obtained using fold change compared to the *t*-Test to rank genes. It should be noted that fold change was used to rank the combined data set but it was compared to the *t*-Test applied to rank the *7.2sox10*:GFP data set and this may account for some of the differences.

As two separate transgenic lines were evaluated under the same experimental conditions, two independent sets of microarray data were available for comparison. Genes that were designated as differentially regulated in both data sets were examined (**Table 57** and **Table 58**). A large number of genes known to be NC or otic vesicle expressed and or validated as down-regulated during the *in situ* hybridization screen were identified by this method. In comparison to the other methods employed, this gene list had the highest true positive hit rate and the lowest false positive hit rate as far as could be determined. False positive genes have been designated as differentially regulated by chance and would therefore not be expected to appear in both data sets. Thus the genes identified by

this method represent attractive targets for future examination by *in situ* hybridization. A down side to this method is that a number of genes validated during the *in situ* hybridization screen were not designated as down-regulated in the *4.9sox10*:GFP data set. Thus this comparison method will exclude a number of genes that are known to be differentially regulated in *sox10* mutant embryos. This method will only be as reliable as the least reliable data set. To overcome this, the number of arrays hybridized with RNA from each transgenic line could be increased. By increasing array number the number of data points for each probe set is increased thereby improving the reliability of any statistical test performed on each probe set. Identifying genes identified as differentially regulated in two independent data sets would then potentially be a very effective way to eliminate false positive results.

During the analysis of microarray data the chosen statistical test is carried out on each gene, leading to thousands of tests being performed. During this process it is likely that a large number of false positives will accumulate (Cui and Churchill, 2003). To address this problem, methods to control for multiple testing can be applied. There are two commonly used methods FWER and FDR. FWER criteria are very strict and designed to eliminate false positive results (Cui and Churchill, 2003). This can result in the exclusion of truly differentially regulated genes; the power of the test is reduced. During the process of a microarray screen a proportion of false positives may be accepted to ensure that a large number of differentially regulated genes can be detected. The FDR or false discovery rate is the proportion of false positives amongst all the genes initially identified as differentially regulated (Cui and Churchill, 2003). The FDR provides a measure of the proportion of false positives that are likely to be present in a data set. As such the FDR method of controlling for multiple testing may be more appropriate. By allowing for a higher rate of false positives to occur this is a more inclusive test that can achieve more power than the FWER method and therefore designate more genes as differentially regulated. The FDR method was applied to the combined data set and the genes detected as differentially regulated were examined (**Table 59** and **Table 60**). 32 genes were designated as down-regulated at $p<0.05$ (5 % of genes are predicted to be false positives). Of these genes, 26 had been assessed during the *in situ* hybridization screen and 15 showed NC or otic vesicle expression. Thus 11/26 genes were false positives (42 %) which was considerably higher than expected. Thus while this method when applied to the combined data set increased the positive hit rate and reduced the false positive hit rate obtained during the *7.2sox10*:GFP validation screen, this method was not as successful as expected. It is unclear why the false positive hit rate was much higher than 5 %, but multiple testing correction methods only identified differentially regulated genes in the combined data set. When FDR methods were applied to the two independent data sets, no genes were identified as differentially regulated (see Chapter

3). Thus an increase in sample size may improve the results of such methods. It may be that this result also highlights the gap between the theoretical result of a statistical test and the actual performance of a test when applied to biological data evaluated by experimental techniques. It should be noted that a large number of biologically relevant results were obtained by screening the *7.2sox10*:GFP data set in which no genes were designated as differentially regulated after controlling for multiple testing. Thus while correcting for multiple testing is the accepted best practise (Cui and Churchill, 2003) in analysing microarray data, successful results can be generated when these controls are not applied.

While no method of ranking genes applied during this project could eliminate false positives, all methods did identify a large number of validated genes. Thus all methods could have been used to select a number of genes for a successful *in situ* hybridization screen. The method chosen and its implementation would depend on the nature of the subsequent *in situ* hybridization validation screen. If the investigator has access to a cDNA library and perhaps an automated *in situ* hybridization machine then a high through put screen is possible. The investigator could then select a large number of genes for validation by relaxing arbitrary criteria for selecting a method that identifies a large number of genes as differentially regulated. This would be done with the aim of validating the greatest number of novel targets and known targets possible. Similarly, if targets were to be validated through a PCR approach then a greater number of genes could be investigated. An investigator lacking access to these resources might consider reducing the number of genes requiring validation using published data, all genes with a clear expression pattern lacking expression in the cell type or tissue of interest could be excluded from the *in situ* hybridization screen.

Published studies that perform a microarray based screen do not publish data comparing different methods of identifying or ranking differentially regulated genes. Perhaps because groups that perform these experiments and validate the biological relevance of the microarray data are not experts in the mechanics of different statistical tests. Equally, when different statistical tests are compared on published data sets by experts in microarray analysis, for example the *t*-Test and the CLEAR test (Valls et al., 2008), the biological relevance of the different results obtained by each method are not investigated. Thus while according to statistical measures the CLEAR test may perform best in identifying supposed differentially regulated genes, no validation of this is performed via methods such as QRT-PCR or *in situ* hybridization. To some extent this chapter has attempted to assess the effects of different methods that can be applied to identify differentially regulated genes by combining *in situ* hybridization data from this project and published data on ZFIN. However, such an approach cannot allow fair comparisons to be made as well annotated and characterised genes are always

preferentially represented in the published literature while poorly annotated genes investigated during a blind *in situ* hybridization screen often display unclear expression patterns. As such the chief focus of this chapter was not to compare different methods but to use them to draw out key genes of interest for future validation.

5.3.4 Selecting Genes by Function.

An aim of microarray expression studies can be to understand the biological functions that are altered between experimental conditions to gain a global view of the biological processes or pathways that are differentially regulated (Al-Shahrour et al., 2008). The web based tool Babelomics can, for example, evaluate if certain GO terms are significantly enriched in the genes identified as up or down-regulated when compared to genes whose expression level is not significantly changed. This could indicate if genes that play a role in, for example, apoptosis or cell division are enriched amongst the differentially regulated genes. This type of analysis could be misleading when a large number of genes designated as differentially regulated are false positives. For this reason, this type of analysis performed on the microarray data sets generated during this project was not presented here. It should be noted that this analysis did highlight the large number of vATPase and IMP biosynthesis pathway genes that were identified as differentially regulated, something that was also shown during the *in situ* hybridization screen.

To identify genes by function, all up or down-regulated genes were categorised by GO term; GO biological process and GO molecular function terms were examined. From these identified GO terms, terms likely to contain transcription factor genes were selected. It proved difficult to select a GO term that covered all transcription factor genes. For example, the GO Term “Transcription factor activity” represented a number of transcription factors but not the key melanocyte transcription factor *mitfa*. Thus the selection of genes by this method is dependent on accurate annotation. This was complicated by having two or more similar GO terms at the same level that, for example GO molecular function level 5 “Transcription factor activity” and “Transcriptional activator”. These two terms covered slightly different sets of genes. GO terms are organised into levels with more specific functional information at higher numbered levels and more general GO terms contained in the lower levels. Thus there is a balance between selecting a very general GO term that includes all transcription factors and a large number of genes that are not transcription factors and selecting a more specific GO term that actually excludes some transcription factors. In addition more specific GO terms are only applied to well characterised genes so some genes of interest are excluded by examining a more specific GO term. Indeed a large number of the differentially regulated genes had no associated GO term. Therefore GO term analysis will favour genes that are well

characterised. All of these factors made GO term analyses very difficult to apply and then interpret, even without considering that a number of the identified genes were likely to be false positives. Therefore even when trying to address a very specific biological question by examining transcription factors, GO term analyses proved, in this case, to not be very informative. However other types of types of target could be investigated through GO terms, for example genes with “receptor activity” GO term: GO:0004872. Other strategies could be considered for the functional analysis of microarray data. Genes of interest could be selected manually based on name, gene family, a literature search or functional annotation. Such manual strategies would be time consuming but would enable the investigator to make informed choices. Essentially, all these methods could be used to interrogate existing microarray data to address a specific biological question of interest with the best approach likely to vary from case to case.

5.3.5 Interesting candidates for future validation

While all the validation of all the genes highlighted during this chapter is of interest, those genes that have been shown to be NC and otic vesicle expressed are of particular interest. The work presented in this chapter has also highlighted some additional genes that have no clear published expression pattern that are also of interest and should be prioritised for validation; these genes are summarised in **Table 64**. For the down-regulated genes, *bhlhb3l* is of interest as this gene was identified as differentially regulated by all three different methods applied in this chapter, in addition this gene is a TF. *Bhlhb3l* expression was assessed during the *in situ* hybridization screen and had no clear expression pattern but the results presented here suggest this gene is worth revisiting. Two other genes are of interest as they were identified as differentially regulated by two methods, *wu:fc64c05* and *atp6v1g1*. The *wu:fc64c05* gene has no published expression pattern thus elucidating this is of interest. *Atp6v1g1* along with *atp6v1ba* and *slc3a2* are of interest as they display potential NC expression patterns on ZFIN, in addition similar genes to these were validated during the *in situ* hybridization screen. The *bhlhb2* gene has been highlighted as this gene is not only a TF but is also NC expressed, according to ZFIN, albeit restricted to an earlier stage than was assessed by this microarray project. *Bhlhb2* was examined during the *in situ* hybridization screen but may also be worth revisiting.

For the up-regulated genes, all of the genes presented in **Table 64** were highlighted as differentially regulated by at least one of the methods utilised during this chapter. The genes *mpv17, meis3* and *sema3d* have all been highlighted as these genes have been annotated as NC expressed on ZFIN thus validating these genes may help to identify more NC expressed genes up-regulated in *sox10* mutant embryos. The genes

syt4, vamp1 and *zic2b* are all expressed in the dorsal NT according to ZFIN and as such may be expressed in NCCs prior to delamination. The *Zic* family of genes are considered to be a constituent of the network of genes that specify early NCCs (Steventon et al., 2005). The genes *ngfr* and *si:dkey-73n10.1* were selected as these were highlighted by two of the methods used during this chapter and have no published expression pattern on ZFIN. In addition, *ngfr* was selected as according to the model of Sox10 function presented in Chapter 1, NC expressed receptors are strong candidates to be up-regulated in *sox10* mutant embryos.

In addition to the genes prioritised for validation that are presented in **Table 64**, a large number of NC and otic vesicle expressed genes that were not examined during the *in situ* hybridization screen have been highlighted in this chapter. These genes also represent strong candidates for future validation studies.

Down-Regulated				
Affymetrix ID	Gene Name	Fold Change	Differentially expressed in 7.2 and 4.9 data sets	FDR *p*-value
Dr.25257.1.A1_at	*bhlhb3l*	-8.42	✓	0.0028
Dr.2268.1.A1_at	*wu:fc64c05*	-2.13	✓	
Dr.5513.1.S1_at	*slc3a2*		✓	
Dr.186.1.S1_at	*atp6v1ba*		✓	
Dr.370.1.S1_at	*atp6v1g1*		✓	0.0251
Dr.3180.1.A1_at	*bhlhb2*			
Up-Regulated				
Affymetrix ID	Gene Name	Fold Change	Differentially expressed in 7.2 and 4.9 data sets	FDR *p*-value
Dr.17843.1.A1_at	*mpv17*	+3.14		
Dr.8061.1.S1_at	*sema3d*	+2.16		
Dr.1680.1.S1_at	*meis3*		✓	
Dr.13868.1.S1_at	*syt4*	+2.04	✓	
Dr.25500.1.S1_at	*vamp1*		✓	
Dr.10689.1.S1_at	*zic2b*		✓	
Dr.10488.1.A1_at	*ngfr*	+2.28	✓	
Dr.6063.1.A1_at	*si:dkey-73n10.1*		✓	0.0054

Table 64: Summary of genes to prioritise for validation.
Genes in this table have been identified as priorities for validation by the methods applied during this chapter.

5.3.6 Regarding the combined dataset

The combined data set was generated in order to appreciate what effect an increase in array number would have on the experiment, particularly on the results of key differentially

regulated marker genes. This data set was unique in that genes were still designated as differentially regulated after controlling for multiple testing. This gave the first indication of a positive effect. Examining the *p*-values of a range of key NC marker genes across all three data sets also indicated that the combined data set performed the best at identifying these genes as differentially regulated (**Table 48**). The combined data set designated almost all of the NC expressed genes validated during the *in situ* hybridization as differentially regulated (**Table 47**). This held true for NC expressed genes but the performance of otic vesicle expressed genes in this data set was unclear due to the paucity of validated marker genes.

Despite NC marker genes being consistently identified as differentially regulated in the combined data set, the question remained if this was a valid experimental data set. The data suggested that otic vesicle expressed genes were not reliably designated as differentially regulated in this data set. While the key marker gene *otomp* was down-regulated in the combined data set, the strongly down-regulated gene *cldnj* was not identified as differentially regulated (**Table 53**). This was the consequence of combining two data sets in which this gene was designated as down-regulated in one and not significantly changed in the other. The unreliable designation of ear genes in the combined data set stems from combining data from two different transgenic lines that label the otic vesicle differently. The *7.2sox10*:GFP line labels the otic vesicle cells strongly with GFP while the in *4.9sox10*:GFP line the otic vesicle cells are weakly labelled with GFP. As such it was expected that otic vesicle expressed genes would not be identified as differentially regulated in the *4.9sox10*:GFP samples because fewer otic vesicle cells would be sorted by FACS in these samples. The consequence of this would be an apparent reduction of otic vesicle gene expression in the *4.9sox10*:GFP samples in comparison to the *7.2sox10*:GFP samples. This variability across samples when the data sets were combined was likely to result in these genes not being identified as differentially regulated. The *t*-Test will not designate a gene as differentially regulated if it displays variability in gene expression across samples. Therefore it seemed the best data set to identify otic vesicle expressed genes in was the *7.2sox10*:GFP data set. It has not been possible to examine enough otic vesicle markers to fully confirm this. However both transgenic lines label NCCs in an almost identical fashion (Carney, 2003, Carney et al., 2006, Dutton et al., 2008). Therefore for NCCs both transgenic lines were effectively equivalent so combining data sets simply increases the number of replicates. This was supported by the results of the key NC marker genes in this data set. To back this up, the greatest number of ZFIN annotated NC expressed genes were identified as differentially regulated in the combined data set. Therefore, for NC expressed genes this data set provides the most reliable list of differentially regulated genes to interrogate with specific biological questions and to source genes for validation from.

Chapter 6: Final Discussion

The initially multipotent cells of the neural crest produce a wide range of different derivatives that migrate to various locations within the developing embryo. Thus this cell population undergoes a number of key processes including cell specification, cell differentiation and the acquisition of migratory properties. These processes have been shown to fail in zebrafish *sox10* mutant embryos as a result of the primary failure in non-skeletogenic NCC specification (Dutton et al., 2001a). Thus the TF Sox10 directly or indirectly regulates a large number of cellular processes. The molecular players and correlates of these processes are poorly understood therefore this project was undertaken in an attempt to identify these genes. The rationale behind this experiment is that direct and indirect targets of Sox10 will show differential expression in *sox10* mutant embryos compared to WT embryos. Thus a comparison of the transcriptome of WT and *sox10* mutant embryos would identify genes whose expression is enhanced by Sox10 function (down-regulated in mutant embryos) or inhibited by Sox10 function (up-regulated in mutant embryos). To specifically target the transcriptome of cells in which *sox10* is expressed, *sox10* expressing cells were enriched, by FACS, from *sox10*:GFP transgenic zebrafish embryos. This was performed on 24 hpf embryos as NC derivatives are undergoing specification, differentiation and migration at this developmental stage. The cell populations that express *sox10* are the neural crest and the otic epithelium. Cells were enriched from both WT and *sox10* mutant embryos and RNA purified from these cells was hybridized to Affymetrix zebrafish microarrays (by the UCI DNA and Protein Microarray Facility, CA, USA). By comparing WT and mutant RNA hybridized microarrays, genes that were differentially regulated were identified. Subsequently these candidates were examined by *in situ* hybridization to validate differential gene expression in the cell types of interest. 27 individual genes were identified and validated in this way.

The *in situ* hybridization screen presented in Chapter 4 validated 23 NC and 3 otic vesicle expressed down-regulated genes and 2 up-regulated otic vesicle expressed genes. The majority of the NC genes were expressed in pigment cells, particularly the xanthophore and melanocyte lineages. This strongly suggests that at 24 hpf, NCCs are chiefly undergoing specification to the pigment cell lineages. Of these validated genes, 9 had not previously been shown to be NC or otic vesicle expressed and 17 had no identified functional link to NC or otic vesicle biology. Thus this project has identified novel marker genes for NCC lineages, genes that previously had no functional association with NC or otic vesicle biology and marker genes known to play roles in NC and otic vesicle biology. This presents a number of opportunities to evaluate the function of these novel genes, chiefly through morpholino mediated gene knock down. Such functional studies could open new avenues of research for NC and otic vesicle biology. In particular, no

function for a connexin gene in early NCCs has been described yet, thus the function of *cx33.8* in the NC is of great interest.

Sox10 functions to specifiy non-skeletogenic NC derivatives, the failure of NCCs to differentiate, migrate and their subsequent apoptosis in *sox10* mutant embryos are secondary defects (Dutton et al., 2001a). Despite the majority of genes having no known functional link to NC biology, it is important to consider how the validated genes identified are likely to fit into this model of Sox10 function. The only validated gene known to be involved in the specification of a specific cell lineage was *mitfa*, the melanocyte lineage master switch TF (Elworthy et al., 2003). Several of the known marker genes validated during the *in situ* hybridization screen are involved in the differentiation of NCC lineages, these include the TF Pax7 (Minchin and Hughes, 2008) and the enzymes Dct and Pah (Ludwig et al., 2004, Schallreuter et al., 2008). The TF Foxd3 is down-regulated in the glia of *sox10* mutant embryos, but the role of Foxd3 in these cells may be to promote cell survival but this is unclear (Stewart et al., 2006). Despite the lack of functional information pertaining to the remaining genes, possible functions in the NC can be hypothesised based on the expression pattern of the gene and current published literature regarding their function in other cell types. The enzymes Atic, Paics and Adsl are all involved in the IMP biosynthesis pathway. It is likely this pathway supplies GTP as a substrate for xanthophore pigmentation therefore these enzymes are involved in differentiation. The v-ATPase subunit genes may play a role in melanosome development and therefore melanocyte differentiation. The enzyme degs1 may be involved in cell migration or cell proliferation (Cadena et al., 1997, Zhou et al., 2009, García-López et al., 2005). The putative tetraspanin zgc:100919, coro1c and rbp4l may all be involved in cell migration (García-López et al., 2005, Rosentreter et al., 2007, Matsukawa et al., 2004). The published data that is available for several genes has no obvious connection to NC biology, for example aldh2b is an enzyme involved in aldehyde metabolism, *slc2a15b* is likely to encode a sugar transporter of some kind, zgc:110343 may have a role as an antioxidant and zgc:110239 may function in proteolysis. For other genes, Dr.4612.1.A1_at and Dr.3972.1.S1_at, a lack of functional information means it is impossible to speculate on the role of these genes in the NC. It would be intriguing to identify the function of all the genes validated during the *in situ* hybridization screen in the NC to further confirm the function of Sox10 as an activator of key genes involved in NC biology. A wide range of different genes have been identified as down-regulated in *sox10* mutant NCCs. It seems likely that these genes function in a number of different roles. The simplest explanation for this is that Sox10 functions to specify non-skeletogenic NCCs and when this fails, a large number of cellular processes are not initiated.

The model of Sox10 function put forward in Chapter 1 includes a prediction of the properties of NCCs in *sox10* mutant embryos. These cells cannot specify correctly and as

such might maintain the properties of early NCCs, perhaps even NCSCs. Therefore genes that mark an intermediate cell type, for example *ltk* marking chromatoblasts (Lopes et al., 2008), markers of NCSCs and receptors for extrinsic signals required to promote NC fate specification might be up-regulated in *sox10* mutant NCCs. NCSCs have not yet been identified in zebrafish therefore any markers of these cells would be extremely valuable. As no such targets were identified during this study, the validity of this part of the model remains untested. The unspecified mutant cell population might express a range of different molecular markers, perhaps for short periods of time. This chaotic molecular situation would ensure these genes could not be detected as up-regulated by microarray analysis. However, *ltk* is strongly up-regulated in *sox10* mutant NCCs arguing against this hypothesis and suggesting that other up-regulated targets do exist.

Down-regulated and up-regulated otic vesicle expressed genes were identified in *sox10* mutant embryos during this project. This included *otomp* which was already known to be down-regulated in *sox10* mutant embryos. A number of otic vesicle expressed genes known to be up-regulated in *sox10* mutant embryos such as *pax2a* and *fgf8* (Dutton et al., 2009) were not identified. It was proposed by Dutton et al., 2009 that the loss of Sox10 function in the ear results in a loss of regulation of key patterning genes with gene expression up-regulated in mutants and occurring in ectopic locations. Without functional information for the two up-regulated otic vesicle expressed genes *stc2* and *sfrp5* it is impossible to tell if these two genes fit into this model but their distinct expression patterns in WT embryos certainly support this. Morpholino knock down of the down-regulated gene *cldnj* results in reduced otoliths formation (Hardison et al., 2005). This correlates with the small otoliths seen in *sox10* mutant embryos and loss of *cldnj* expression is likely to contribute to this phenotype. Thus again the otic vesicle expressed genes identified during this study seem to fit with current data.

Throughout this project the performance of a number of key marker genes, for example *mitfa* and *dct*, was assessed. These genes were known to be down-regulated in *sox10* mutant embryos. Given that our group had very little experience with FACS and had no prior experience of microarray experiments, these genes served as markers at vital checkpoints. Expression levels of these key marker genes in purified GFP positive populations of WT and mutant cells were evaluated by RT-PCR to substantiate differential gene expression prior to sending samples for microarray analysis. RT-PCR was also used to assess the expression of these genes between GFP positive and GFP negative cell populations after FACS. Thus without these marker genes the efficacy of the FACS protocol would have been difficult to evaluate. The use of these marker genes also extended into the microarray section of the project. Not only was the performance of these genes in the raw microarray data examined but the performance of these genes was used to evaluate different GEPAS Expresso normalization methods in combination with

different methods applied using the T-Rex tool to identify differentially regulated genes. This led to, in conjunction with published literature, selecting the RMA normalization method and *t*-Test to process and analyse the raw microarray data. Thus throughout the project these marker genes provided results that not only confirmed the success of each step but also enabled important choices to be made. These marker genes were therefore invaluable.

While the project has identified a large number interesting targets of Sox10, the majority of genes (nearly 80 %) examined during the *in situ* hybridization screen were false positives. This high false positive hit rate could not be resolved through changing the method by which differentially regulated genes were identified. Thus a large amount of validation work was performed on false positive microarray results. The false positive hit rate for microarray experiments, when published, seems to vary greatly from study to study. A microarray study was performed to identify genes whose expression was restricted to either deep or superficial ectodermal cells in *Xenopus* embryos (Chalmers et al., 2006). When the differentially expressed genes (twofold cut off) identified by microarray analysis were examined by *in situ* hybridization, 32/98 (33 %) genes were validated. Thus 67 % of genes were false positives. This screen also adopted a blind *in situ* hybridization approach to validate target genes. However, many aspects of this experiment were different; the microarray platform used was different being a two colour array, the model organism and biological question being examined were different, the cells were treated differently prior to RNA extraction and fold change was used to identify differentially expressed genes rather than a *t*-Test. These are only a few selected differences from the project presented here. A study in *Drosophila* used FACS to select proneural cluster cells prior to RNA extraction for hybridization to Affymetrix microarrays (Reeves and Posakony, 2005). Microarray analysis was performed to identify genes expressed preferentially in the proneural cluster cells (twofold increase). This identified 204 candidate genes including 23 genes known to be expressed in the proneural cluster cells. 43 genes not known to be expressed in the cells of interest were selected and 27 (63 %) were validated by *in situ* hybridization. Thus this study purified cells in a similar way to the project presented here but had a false positive hit rate of 37 % despite excluding known proneural cluster expressed genes from their study. This could to some extent stem from only a portion of the differentially regulated genes being selected for validation. Two microarray projects identified genes differentially regulated in *cloche* mutants in comparison to WT embryos (Sumanas et al., 2005, Qian et al., 2005). Thus these studies are interested in examining a similar biological question to this project. Both studies differed in microarray platforms, examining zebrafish embryos at slightly different embryonic stages, differentially expressed genes were defined by fold change in both studies but with different cut off values and both studies used different methods of

validation. Despite investigating the same biological question, only 12 validated genes were common between both screens. Thus differences exist between studies that investigate the same biological question but use slightly different methods. While trying to investigate published literature for true positive and false positive hit rates it became clear that direct comparisons between microarray studies is difficult. Not only are the biological questions and model organisms being studied different but so are the microarray platforms, experimental methods, analytical methods and validation methods. In addition, the exact protocol of each step can vary between labs, even if the same method is used. This can lead to striking differences between the genes validated in quite similar studies (Sumanas et al., 2005, Qian et al., 2005). Therefore while the false positive hit rate of the study presented here does seem to be high, it is difficult to relate this to other studies. The high false positive error rate of this study may result from variation in gene expression in each of the pooled samples of cells from which RNA was extracted. It cannot be determined if this is caused by natural variation between samples or if it is a result of experimental manipulation. It is certainly possible that by performing FACS the extensive manipulation of the embryos and cells prior to RNA extraction has had an effect on gene expression. Alternatively, the use of an RNA amplification step may have had an effect. While it has been shown that microarray data from unamplified and amplified sources of RNA are capable of identifying a large number of the same genes (Qian et al., 2005), it is possible that amplification could increase the false positive hit rate. However, as a number of biologically relevant results were identified in the microarray data and subsequently validated, such manipulations did not affect endogenous gene expression detrimentally.

If this project was to be repeated, a number of factors could be changed. During the FACS purification of GFP positive cells it would be sensible to introduce pulse width to the sorting characteristics to distinguish carefully between single cells and clumps of cells. This would help to ensure only single cells were collected and therefore if GFP positive cells were clumped with GFP negative cells, prevent these GFP negative cells contaminating the samples. By improving the homogeneity of the collected cell populations the microarray data might become more consistent between each array and prevent some false positives being identified as differentially regulated. At the microarray normalization stage it might be worth investigating whether the RMA normalization method which excludes the mismatch probe sets does perform better, as was shown in the literature (Irizarry et al., 2003b), than the Affymetrix normalization method which includes mismatch probe set data. Also it might be worth investigating alternative methods to the *t*-Test to identify differentially regulated genes, such as the CLEAR test as different methods do produce differing results (Millenaar et al., 2006). However, as a good number of true positive genes were validated from the microarray data the methods utilised during this project were effective. The best way to improve the identification of genes as

differentially regulated or not is likely to be by increasing the number of microarrays used. The effects of this was demonstrated by the improved reliability of the data generated from analysing the combined data set (5 WT versus 5 mutant arrays) than the *7.2sox10*:GFP (3 WT versus 3 mutant arrays) or *4.9sox10*:GFP (2 WT versus 2 mutant arrays) data sets.

The *in situ* hybridization screen could also have been streamlined to save both time and resources from investigating false positives. As was examined in Chapter 5, published expression data on the ZFIN website could have been used to initially screen the microarray data *in silico* to eliminate likely false positive genes with clear published expression patterns. This would lead to genes known to be expressed in the cell types of interest and genes with no published expression pattern that are potentially novel being prioritised. Such an approach would certainly benefit a lab that does not have access to a cDNA library to facilitate the generation of *in situ* hybridization probes. In addition, a number of genes from interesting gene families identified as differentially regulated were selected for *in situ* hybridization analysis prior to commencing the unbiased screen presented in Chapter 4. The hope was that investigating a few additional targets quickly using *in situ* probes kindly donated by other labs would validate differential gene expression and increase confidence that the microarray data analysis had been effective. In reality, selecting a few genes did not shed any light on this, particularly given that such a large percentage of genes identified as differentially regulated were false positives. It would have been more beneficial to start the screen and examine the 10 most down-regulated genes immediately. This set of genes contained 4 differentially regulated true positives.

Given that a number of very interesting targets of Sox10 have been identified during this study, the project could be extended. One obvious extension would be to change the time point of the embryos being examined. This study was performed at the earliest point when WT and *sox10* mutant embryos could be distinguished by GFP pattern in *sox10*:GFP transgenic embryos. Thus to extend this study to an earlier time point morpholino injection that knocks down *sox10* expression would be required (Dutton et al., 2001b). Morphant embryo and WT embryo RNA from GFP positive cells could then be compared by microarray analysis. Alternatively the microarray analysis could be extended to a later time point. Given that this project mostly identified pigment cell expressed genes, the study could focus on the change of gene expression in PNS cells in *sox10* mutant embryos by selecting time points, for example 48 hpf, when cells of the enteric nervous system or dorsal root ganglia are being specified. This could help to address the paucity of marker genes for these cell types and help to dissect the role that Sox10 plays in these cell types.

To summarise, the first project of this type to be carried out by our group has identified a large number of genes that are differentially regulated in *sox10* mutant NC and otic vesicle cells. These genes are direct or indirect targets of Sox10. Not only has the identification of these genes opened up a number of new avenues for research but has also unearthed a number of marker genes that will prove useful in future research. Further *in situ* hybridization screening will validate yet more genes that are differentially regulated in *sox10* mutant embryos. The genes discovered by this approach include both known NC and otic vesicle genes and those with no prior association to these cell types. Such identification of novel genes would be extremely difficult to achieve through a different approach. The range of genes that have been validated during the *in situ* hybridization screen appear to fit within the model of Sox10 specifying non-skeletogenic NCCs.

Bibliography

AAKU-SARASTE, E., HELLWIG, A. & HUTTNER, W. 1996. Loss of occludin and functional tight junctions, but not ZO-1, during neural tube closure--remodeling of the neuroepithelium prior to neurogenesis. *Dev Biol,* 180**,** 664-79.

ADAMS, M., GAMMILL, L. & BRONNER-FRASER, M. 2008. Discovery of transcription factors and other candidate regulators of neural crest development. *Dev Dyn,* 237**,** 1021-33.

AFFYMETRIX, I. 2008. *www.Affymetrix.com*.

ALBERTS, B., BRAY, D., JOHNSON, A., LEWIS, J., RAFF, M., ROBERTS, K. & WALTER, P. 1998. *Essential Cell Biology,* New York, Garland Publishing, Inc.

ALFANDARI, D., COUSIN, H., GAULTIER, A., SMITH, K., WHITE, J., DARRIBÈRE, T. & DESIMONE, D. 2001. Xenopus ADAM 13 is a metalloprotease required for cranial neural crest-cell migration. *Curr Biol,* 11**,** 918-30.

AL-SHAHROUR, F., CARBONELL, J., MINGUEZ, P., GOETZ, S., CONESA, A., TÁRRAGA, J., MEDINA, I., ALLOZA, E., MONTANER, D. & DOPAZO, J. 2008. Babelomics: advanced functional profiling of transcriptomics, proteomics and genomics experiments. *Nucleic Acids Res,* 36**,** W341-6.

AL-SHAHROUR, F., MINGUEZ, P., TÁRRAGA, J., MONTANER, D., ALLOZA, E., VAQUERIZAS, J., CONDE, L., BLASCHKE, C., VERA, J. & DOPAZO, J. 2006. BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments. *Nucleic Acids Res,* 34**,** W472-6.

AMSTERDAM, A., NISSEN, R. M., SUN, Z., SWINDELL, E. C., FARRINGTON, S. & HOPKINS, N. 2004. Identification of 315 genes essential for early zebrafish development. *Proc Natl Acad Sci U S A,* 101**,** 12792-7.

ANCANS, J. & THODY, A. 2000. Activation of melanogenesis by vacuolar type H(+)-ATPase inhibitors in amelanotic, tyrosinase positive human and mouse melanoma cells. *FEBS Lett,* 478**,** 57-60.

ANCANS, J., TOBIN, D., HOOGDUIJN, M., SMIT, N., WAKAMATSU, K. & THODY, A. 2001. Melanosomal pH controls rate of melanogenesis, eumelanin/phaeomelanin ratio and melanosome maturation in melanocytes and melanoma cells. *Exp Cell Res,* 268**,** 26-35.

AOKI, Y., SAINT-GERMAIN, N., GYDA, M., MAGNER-FINK, E., LEE, Y. H., CREDIDIO, C. & SAINT-JEANNET, J. P. 2003. Sox10 regulates the development of neural crest-derived melanocytes in Xenopus. *Dev Biol,* 259**,** 19-33.

ASHBURNER, M., BALL, C., BLAKE, J., BOTSTEIN, D., BUTLER, H., CHERRY, J., DAVIS, A., DOLINSKI, K., DWIGHT, S., EPPIG, J., HARRIS, M., HILL, D., ISSEL-TARVER, L., KASARSKIS, A., LEWIS, S., MATESE, J., RICHARDSON, J., RINGWALD, M., RUBIN, G. & SHERLOCK, G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet,* 25**,** 25-9.

AYBAR, M. J. & MAYOR, R. 2002. Early induction of neural crest cells: lessons learned from frog, fish and chick. *Curr Opin Genet Dev,* 12**,** 452-8.

BALDI, P. & LONG, A. 2001. A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes. *Bioinformatics,* 17**,** 509-19.

BANNERMAN, P., NICHOLS, W., PUHALLA, S., OLIVER, T., BERMAN, M. & PLEASURE, D. 2000. Early migratory rat neural crest cells express functional gap junctions: evidence that neural crest cell survival requires gap junction function. *J Neurosci Res,* 61**,** 605-15.

BARRIONUEVO, F. & SCHERER, G. 2009. SOX E genes: SOX9 and SOX8 in mammalian testis development. *Int J Biochem Cell Biol*.

BASCH, M. L., BRONNER-FRASER, M. & GARCIA-CASTRO, M. I. 2006. Specification of the neural crest occurs during gastrulation and requires Pax7. *Nature,* 441**,** 218-22.

BASRUR, V., YANG, F., KUSHIMOTO, T., HIGASHIMOTO, Y., YASUMOTO, K., VALENCIA, J., MULLER, J., VIEIRA, W., WATABE, H., SHABANOWITZ, J., HEARING, V., HUNT, D. & APPELLA, E. 2003. Proteomic analysis of early melanosomes: identification of novel melanosomal proteins. *J Proteome Res,* 2**,** 69-79.

BHATNAGAR, V., ANJAIAH, S., PURI, N., DARSHANAM, B. & RAMAIAH, A. 1993. pH of melanosomes of B 16 murine melanoma is acidic: its physiological importance in the regulation of melanin biosynthesis. *Arch Biochem Biophys,* 307**,** 183-92.

BI, W., DENG, J., ZHANG, Z., BEHRINGER, R. & DE CROMBRUGGHE, B. 1999. Sox9 is required for cartilage formation. *Nat Genet,* 22**,** 85-9.

BOCCALATTE, F., VOENA, C., RIGANTI, C., BOSIA, A., D'AMICO, L., RIERA, L., CHENG, M., RUGGERI, B., JENSEN, O., GOSS, V., LEE, K., NARDONE, J., RUSH, J., POLAKIEWICZ, R., COMB, M., CHIARLE, R. & INGHIRAMI, G. 2009. The enzymatic activity of 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase is enhanced by NPM-ALK: new insights in ALK-mediated pathogenesis and the treatment of ALCL. *Blood,* 113**,** 2776-90.

BOLSTAD, B. M., IRIZARRY, R. A., ASTRAND, M. & SPEED, T. P. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics,* 19**,** 185-93.

BONDURAND, N., DASTOT-LE MOAL, F., STANCHINA, L., COLLOT, N., BARAL, V., MARLIN, S., ATTIE-BITACH, T., GIURGEA, I., SKOPINSKI, L., REARDON, W., TOUTAIN, A., SARDA, P., ECHAIEB, A., LACKMY-PORT-LIS, M., TOURAINE, R., AMIEL, J., GOOSSENS, M. & PINGAULT, V. 2007. Deletions at the SOX10 gene locus cause Waardenburg syndrome types 2 and 4. *Am J Hum Genet,* 81**,** 1169-85.

BONDURAND, N., GIRARD, M., PINGAULT, V., LEMORT, N., DUBOURG, O. & GOOSSENS, M. 2001. Human Connexin 32, a gap junction protein altered in the X-linked form of Charcot-Marie-Tooth disease, is directly regulated by the transcription factor SOX10. *Hum Mol Genet,* 10**,** 2783-95.

BONDURAND, N., KOBETZ, A., PINGAULT, V., LEMORT, N., ENCHA-RAZAVI, F., COULY, G., GOERICH, D. E., WEGNER, M., ABITBOL, M. & GOOSSENS, M. 1998. Expression of the SOX10 gene during human development. *FEBS Lett,* 432**,** 168-72.

BONDURAND, N., PINGAULT, V., GOERICH, D. E., LEMORT, N., SOCK, E., LE CAIGNEC, C., WEGNER, M. & GOOSSENS, M. 2000. Interaction among SOX10, PAX3 and MITF, three genes altered in Waardenburg syndrome. *Hum Mol Genet,* 9**,** 1907-17.

BRITSCH, S., GOERICH, D., RIETHMACHER, D., PEIRANO, R., ROSSNER, M., NAVE, K., BIRCHMEIER, C. & WEGNER, M. 2001. The transcription factor Sox10 is a key regulator of peripheral glial development. *Genes Dev,* 15**,** 66-78.

BRONNER-FRASER, M. & FRASER, S. 1989. Developmental potential of avian trunk neural crest cells in situ. *Neuron,* 3**,** 755-66.

CADENA, D., KURTEN, R. & GILL, G. 1997. The product of the MLD gene is a member of the membrane fatty acid desaturase family: overexpression of MLD inhibits EGF receptor biosynthesis. *Biochemistry,* 36**,** 6960-7.

CALLOW, M., DUDOIT, S., GONG, E., SPEED, T. & RUBIN, E. 2000. Microarray expression profiling identifies genes with altered expression in HDL-deficient mice. *Genome Res,* 10**,** 2022-9.

CANO, A., PÉREZ-MORENO, M., RODRIGO, I., LOCASCIO, A., BLANCO, M., DEL BARRIO, M., PORTILLO, F. & NIETO, M. 2000. The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nat Cell Biol,* 2**,** 76-83.

CARNEY, T. 2003. *Generation of transgenic lines for analysis of neural crest development in zebrafish.* Ph.D, University of Bath.

CARNEY, T. J., DUTTON, K. A., GREENHILL, E., DELFINO-MACHIN, M., DUFOURCQ, P., BLADER, P. & KELSH, R. N. 2006. A direct role for Sox10 in specification of neural crest-derived sensory neurons. *Development,* 133**,** 4619-30.

CERDA, G., HARGRAVE, M. & LEWIS, K. 2009. RNA profiling of FAC-sorted neurons from the developing zebrafish spinal cord. *Dev Dyn,* 238**,** 150-61.

CHABOISSIER, M., KOBAYASHI, A., VIDAL, V., LÜTZKENDORF, S., VAN DE KANT, H., WEGNER, M., DE ROOIJ, D., BEHRINGER, R. & SCHEDL, A. 2004. Functional analysis of Sox8 and Sox9 during sex determination in the mouse. *Development,* 131**,** 1891-901.

CHALMERS, A. D., LACHANI, K., SHIN, Y., SHERWOOD, V., CHO, K. W. & PAPALOPULU, N. 2006. Grainyhead-like 3, a transcription factor identified in a microarray screen, promotes the specification of the superficial layer of the embryonic epidermis. *Mech Dev,* 123**,** 702-18.

CHENG, Y., CHEUNG, M., ABU-ELMAGD, M., ORME, A. & SCOTTING, P. 2000. Chick sox10, a transcription factor expressed in both early neural crest cells and central nervous system. *Brain Res Dev Brain Res,* 121**,** 233-41.

CHENG, Y., LEE, C., BADGE, R., ORME, A. & SCOTTING, P. 2001. Sox8 gene expression identifies immature glial cells in developing cerebellum and cerebellar tumours. *Brain Res Mol Brain Res,* 92**,** 193-200.

CHEUNG, M. & BRISCOE, J. 2003. Neural crest development is regulated by the transcription factor Sox9. *Development,* 130**,** 5681-93.

CHEUNG, M., CHABOISSIER, M., MYNETT, A., HIRST, E., SCHEDL, A. & BRISCOE, J. 2005. The transcriptional control of trunk neural crest induction, survival, and delamination. *Dev Cell,* 8**,** 179-92.

CHRISTOPHERSON, R., LYONS, S. & WILSON, P. 2002. Inhibitors of de novo nucleotide biosynthesis as drugs. *Acc Chem Res,* 35**,** 961-71.

CLÉMENT-ZIZA, M., GENTIEN, D., LYONNET, S., THIERY, J., BESMOND, C. & DECRAENE, C. 2009. Evaluation of methods for amplification of picogram amounts of total RNA for whole genome expression profiling. *BMC Genomics,* 10**,** 246.

COLLAZO, A., BRONNER-FRASER, M. & FRASER, S. 1993. Vital dye labelling of Xenopus laevis trunk neural crest reveals multipotency and novel pathways of migration. *Development,* 118**,** 363-76.

COVASSIN, L., AMIGO, J., SUZUKI, K., TEPLYUK, V., STRAUBHAAR, J. & LAWSON, N. 2006. Global analysis of hematopoietic and vascular endothelial gene expression by tissue specific microarray profiling in zebrafish. *Dev Biol,* 299**,** 551-62.

CRUCIANI, V. & MIKALSEN, S. 2006. The vertebrate connexin family. *Cell Mol Life Sci,* 63**,** 1125-40.

CUI, X. & CHURCHILL, G. 2003. Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol,* 4**,** 210.

DAVID, N. B., SAPEDE, D., SAINT-ETIENNE, L., THISSE, C., THISSE, B., DAMBLY-CHAUDIERE, C., ROSA, F. M. & GHYSEN, A. 2002. Molecular basis of cell migration in the fish lateral line: role of the chemokine receptor CXCR4 and of its ligand, SDF1. *Proc Natl Acad Sci U S A,* 99**,** 16297-302.

DE BELLARD, M., RAO, Y. & BRONNER-FRASER, M. 2003. Dual function of Slit2 in repulsion and enhanced migration of trunk, but not vagal, neural crest cells. *J Cell Biol,* 162**,** 269-79.

DE CALISTO, J., ARAYA, C., MARCHANT, L., RIAZ, C. & MAYOR, R. 2005. Essential role of non-canonical Wnt signalling in neural crest migration. *Development,* 132**,** 2587-97.

DELFINO-MACHÍN, M., CHIPPERFIELD, T., RODRIGUES, F. & KELSH, R. 2007. The proliferating field of neural crest stem cells. *Dev Dyn,* 236**,** 3242-54.

DORSKY, R. I., MOON, R. T. & RAIBLE, D. W. 1998. Control of neural crest cell fate by the Wnt signalling pathway. *Nature,* 396**,** 370-3.

DORSKY, R., RAIBLE, D. & MOON, R. 2000. Direct regulation of nacre, a zebrafish MITF homolog required for pigment cell formation, by the Wnt pathway. *Genes Dev,* 14**,** 158-62.

DRIEVER, W., SOLNICA-KREZEL, L., SCHIER, A., NEUHAUSS, S., MALICKI, J., STEMPLE, D., STAINIER, D., ZWARTKRUIS, F., ABDELILAH, S., RANGINI, Z., BELAK, J. & BOGGS, C. 1996. A genetic screen for mutations affecting embryogenesis in zebrafish. *Development,* 123**,** 37-46.

DUONG, T. & ERICKSON, C. 2004. MMP-2 plays an essential role in producing epithelial-mesenchymal transformations in the avian embryo. *Dev Dyn,* 229**,** 42-53.

DUTTON, J., ANTONELLIS, A., CARNEY, T., RODRIGUES, F., PAVAN, W., WARD, A. & KELSH, R. 2008. An evolutionarily conserved intronic region controls the spatiotemporal expression of the transcription factor Sox10. *BMC Dev Biol,* 8**,** 105.

DUTTON, K. A., PAULINY, A., LOPES, S. S., ELWORTHY, S., CARNEY, T. J., RAUCH, J., GEISLER, R., HAFFTER, P. & KELSH, R. N. 2001a. Zebrafish colourless encodes sox10 and specifies non-ectomesenchymal neural crest fates. *Development,* 128**,** 4113-25.

DUTTON, K., ABBAS, L., SPENCER, J., BRANNON, C., MOWBRAY, C., NIKAIDO, M., KELSH, R. & WHITFIELD, T. 2009. A zebrafish model for Waardenburg syndrome type IV reveals diverse roles for Sox10 in the otic vesicle. *Dis Model Mech,* 2**,** 68-83.

DUTTON, K., DUTTON, J. R., PAULINY, A. & KELSH, R. N. 2001b. A morpholino phenocopy of the colourless mutant. *Genesis,* 30**,** 188-9.

EDERY, P., ATTIÉ, T., AMIEL, J., PELET, A., ENG, C., HOFSTRA, R., MARTELLI, H., BIDAUD, C., MUNNICH, A. & LYONNET, S. 1996. Mutation of the endothelin-3 gene in the

Waardenburg-Hirschsprung disease (Shah-Waardenburg syndrome). *Nat Genet,* 12**,** 442-4.
EISEN, J. & WESTON, J. 1993. Development of the neural crest in the zebrafish. *Dev Biol,* 159**,** 50-9.
ELWORTHY, S., LISTER, J. A., CARNEY, T. J., RAIBLE, D. W. & KELSH, R. N. 2003. Transcriptional regulation of mitfa accounts for the sox10 requirement in zebrafish melanophore development. *Development,* 130**,** 2809-18.
ELWORTHY, S., PINTO, J. P., PETTIFER, A., CANCELA, M. L. & KELSH, R. N. 2005. Phox2b function in the enteric nervous system is conserved in zebrafish and is sox10-dependent. *Mech Dev,* 122**,** 659-69.
FAN, L., MOON, J., WONG, T., CRODIAN, J. & COLLODI, P. 2008. Zebrafish primordial germ cell cultures derived from vasa::RFP transgenic embryos. *Stem Cells Dev,* 17**,** 585-97.
FINN, R. & KAPOOR, B. 2008. *Fish Larval Physiology,* USA, Science Publishers.
GAMMILL, L. & BRONNER-FRASER, M. 2002. Genomic analysis of neural crest induction. *Development,* 129**,** 5731-41.
GAMMILL, L., GONZALEZ, C., GU, C. & BRONNER-FRASER, M. 2006. Guidance of trunk neural crest migration requires neuropilin 2/semaphorin 3F signaling. *Development,* 133**,** 99-106.
GARCÍA-LÓPEZ, M., BARREIRO, O., GARCÍA-DÍEZ, A., SÁNCHEZ-MADRID, F. & PEÑAS, P. 2005. Role of tetraspanins CD9 and CD151 in primary melanocyte motility. *J Invest Dermatol,* 125**,** 1001-9.
GARRIOCK, R. J. & KRIEG, P. A. 2007. Wnt11-R signaling regulates a calcium sensitive EMT event essential for dorsal fin development of Xenopus. *Dev Biol,* 304**,** 127-40.
GAUTIER, L., COPE, L., BOLSTAD, B. & IRIZARRY, R. 2004. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics,* 20**,** 307-15.
GELINEAU-VAN WAES, J., SMITH, L., VAN WAES, M., WILBERDING, J., EUDY, J., BAUER, L. & MADDOX, J. 2008. Altered expression of the iron transporter Nramp1 (Slc11a1) during fetal development of the retinal pigment epithelium in microphthalmia-associated transcription factor Mitf(mi) and Mitf(vitiligo) mouse mutants. *Exp Eye Res,* 86**,** 419-33.
GENTLEMAN, R., CAREY, V., BATES, D., BOLSTAD, B., DETTLING, M., DUDOIT, S., ELLIS, B., GAUTIER, L., GE, Y., GENTRY, J., HORNIK, K., HOTHORN, T., HUBER, W., IACUS, S., IRIZARRY, R., LEISCH, F., LI, C., MAECHLER, M., ROSSINI, A., SAWITZKI, G., SMITH, C., SMYTH, G., TIERNEY, L., YANG, J. & ZHANG, J. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol,* 5**,** R80.
GOODENOUGH, D. & PAUL, D. 2003. Beyond the gap: functions of unpaired connexon channels. *Nat Rev Mol Cell Biol,* 4**,** 285-94.
GRABHER, C. & WITTBRODT, J. 2008. Recent advances in meganuclease-and transposon-mediated transgenesis of medaka and zebrafish. *Methods Mol Biol,* 461**,** 521-39.
GREENHILL, E. 2008. *Genetic regulation of neural crest cell differentiation.* Ph.D, University of Bath.
GROVES, J., HAMMOND, C. & HUGHES, S. 2005. Fgf8 drives myogenic progression of a novel lateral fast muscle fibre population in zebrafish. *Development,* 132**,** 4211-22.
GUILLEMOT, F., LO, L., JOHNSON, J., AUERBACH, A., ANDERSON, D. & JOYNER, A. 1993. Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons. *Cell,* 75**,** 463-76.
HAFFTER, P., GRANATO, M., BRAND, M., MULLINS, M., HAMMERSCHMIDT, M., KANE, D., ODENTHAL, J., VAN EEDEN, F., JIANG, Y., HEISENBERG, C., KELSH, R., FURUTANI-SEIKI, M., VOGELSANG, E., BEUCHLE, D., SCHACH, U., FABIAN, C. & NÜSSLEIN-VOLHARD, C. 1996. The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio. *Development,* 123**,** 1-36.
HALBLEIB, J. & NELSON, W. 2006. Cadherins in development: cell adhesion, sorting, and tissue morphogenesis. *Genes Dev,* 20**,** 3199-214.
HAMMOND, C., HINITS, Y., OSBORN, D., MINCHIN, J., TETTAMANTI, G. & HUGHES, S. 2007. Signals and myogenic regulatory factors restrict pax3 and pax7 expression to dermomyotome-like tissue in zebrafish. *Dev Biol,* 302**,** 504-21.
HARDISON, A. L., LICHTEN, L., BANERJEE-BASU, S., BECKER, T. S. & BURGESS, S. M. 2005. The zebrafish gene claudinj is essential for normal ear function and important for the formation of the otoliths. *Mech Dev,* 122**,** 949-58.
HARRINGTON, C., ROSENOW, C. & RETIEF, J. 2000. Monitoring gene expression using DNA microarrays. *Curr Opin Microbiol,* 3**,** 285-91.
HEMLER, M. 2005. Tetraspanin functions and associated microdomains. *Nat Rev Mol Cell Biol,* 6**,** 801-11.
HENION, P. & WESTON, J. 1997. Timing and pattern of cell fate restrictions in the neural crest lineage. *Development,* 124**,** 4351-9.

HENION, P., RAIBLE, D., BEATTIE, C., STOESSER, K., WESTON, J. & EISEN, J. 1996. Screen for mutations affecting development of Zebrafish neural crest. *Dev Genet,* 18**,** 11-7.

HERBARTH, B., PINGAULT, V., BONDURAND, N., KUHLBRODT, K., HERMANS-BORGMEYER, I., PULITI, A., LEMORT, N., GOOSSENS, M. & WEGNER, M. 1998. Mutation of the Sry-related Sox10 gene in Dominant megacolon, a mouse model for human Hirschsprung disease. *Proc Natl Acad Sci U S A,* 95**,** 5161-5.

HOEK, K., SCHLEGEL, N., EICHHOFF, O., WIDMER, D., PRAETORIUS, C., EINARSSON, S., VALGEIRSDOTTIR, S., BERGSTEINSDOTTIR, K., SCHEPSKY, A., DUMMER, R. & STEINGRIMSSON, E. 2008. Novel MITF targets identified using a two-step DNA microarray strategy. *Pigment Cell Melanoma Res,* 21**,** 665-76.

HOFFMAN, T., JAVIER, A., CAMPEAU, S., KNIGHT, R. & SCHILLING, T. 2007. Tfap2 transcription factors in zebrafish neural crest development and ectodermal evolution. *J Exp Zoolog B Mol Dev Evol,* 308**,** 679-91.

HOFSTRA, R., OSINGA, J., TAN-SINDHUNATA, G., WU, Y., KAMSTEEG, E., STULP, R., VAN RAVENSWAAIJ-ARTS, C., MAJOOR-KRAKAUER, D., ANGRIST, M., CHAKRAVARTI, A., MEIJERS, C. & BUYS, C. 1996. A homozygous mutation in the endothelin-3 gene associated with a combined Waardenburg type 2 and Hirschsprung phenotype (Shah-Waardenburg syndrome). *Nat Genet,* 12**,** 445-7.

HONORE, S. M., AYBAR, M. J. & MAYOR, R. 2003. Sox10 is required for the early development of the prospective neural crest in Xenopus embryos. *Dev Biol,* 260**,** 79-96.

HOTH, C., MILUNSKY, A., LIPSKY, N., SHEFFER, R., CLARREN, S. & BALDWIN, C. 1993. Mutations in the paired domain of the human PAX3 gene cause Klein-Waardenburg syndrome (WS-III) as well as Waardenburg syndrome type I (WS-I). *Am J Hum Genet,* 52**,** 455-62.

HU, Y. F., ZHANG, Z.-J. & SIEBER-BLUM, M. 2006. An Epidermal Neural Crest Stem Cell (EPI-NCSC) Molecular Signature. *Stem Cells***,** 2006-0233.

HUANG, G., WESSELS, A., SMITH, B., LINASK, K., EWART, J. & LO, C. 1998. Alteration in connexin 43 gap junction gene dosage impairs conotruncal heart development. *Dev Biol,* 198**,** 32-44.

IIZAKA, M., HAN, H., AKASHI, H., FURUKAWA, Y., NAKAJIMA, Y., SUGANO, S., OGAWA, M. & NAKAMURA, Y. 2000. Isolation and chromosomal assignment of a novel human gene, CORO1C, homologous to coronin-like actin-binding proteins. *Cytogenet Cell Genet,* 88**,** 221-4.

IKENOUCHI, J., MATSUDA, M., FURUSE, M. & TSUKITA, S. 2003. Regulation of tight junctions during the epithelium-mesenchyme transition: direct repression of the gene expression of claudins/occludin by Snail. *J Cell Sci,* 116**,** 1959-67.

IMMENSCHUH, S. & BAUMGART-VOGT, E. 2005. Peroxiredoxins, oxidative stress, and cell proliferation. *Antioxid Redox Signal,* 7**,** 768-77.

INOUE, K., KHAJAVI, M., OHYAMA, T., HIRABAYASHI, S., WILSON, J., REGGIN, J., MANCIAS, P., BUTLER, I., WILKINSON, M., WEGNER, M. & LUPSKI, J. 2004. Molecular mechanism for distinct neurological phenotypes conveyed by allelic truncating mutations. *Nat Genet,* 36**,** 361-9.

INOUE, K., SHILO, K., BOERKOEL, C., CROWE, C., SAWADY, J., LUPSKI, J. & AGAMANOLIS, D. 2002. Congenital hypomyelinating neuropathy, central dysmyelination, and Waardenburg-Hirschsprung disease: phenotypes linked by SOX10 mutation. *Ann Neurol,* 52**,** 836-42.

IRIZARRY, R. A., HOBBS, B., COLLIN, F., BEAZER-BARCLAY, Y. D., ANTONELLIS, K. J., SCHERF, U. & SPEED, T. P. 2003b. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics,* 4**,** 249-64.

IRIZARRY, R., BOLSTAD, B., COLLIN, F., COPE, L., HOBBS, B. & SPEED, T. 2003a. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res,* 31**,** e15.

ISHIBASHI, K. & IMAI, M. 2002. Prospect of a stanniocalcin endocrine/paracrine system in mammals. *Am J Physiol Renal Physiol,* 282**,** F367-75.

ITO, Y., WIESE, S., FUNK, N., CHITTKA, A., ROSSOLL, W., BOMMEL, H., WATABE, K., WEGNER, M. & SENDTNER, M. 2006. Sox10 regulates ciliary neurotrophic factor gene expression in Schwann cells. *Proc Natl Acad Sci U S A,* 103**,** 7871-6.

IWASHITA, T., KRUGER, G., PARDAL, R., KIEL, M. & MORRISON, S. 2003. Hirschsprung disease is linked to defects in neural crest stem cell function. *Science,* 301**,** 972-6.

JIA, L., CHENG, L. & RAPER, J. 2005. Slit/Robo signaling is necessary to confine early neural crest cells to the ventral migratory pathway in the trunk. *Dev Biol,* 282**,** 411-21.

JIAO, Z., MOLLAAGHABABA, R., PAVAN, W. J., ANTONELLIS, A., GREEN, E. D. & HORNYAK, T. J. 2004. Direct interaction of Sox10 with the promoter of murine Dopachrome Tautomerase (Dct) and synergistic activation of Dct expression with Mitf. *Pigment Cell Res,* 17**,** 352-62.

JOHNSON, S. L., AFRICA, D., WALKER, C. & WESTON, J. A. 1995. Genetic control of adult pigment stripe development in zebrafish. *Dev Biol,* 167**,** 27-33.
KANG, S., CHAE, H., SEO, M., KIM, K., BAINES, I. & RHEE, S. 1998. Mammalian peroxiredoxin isoforms can reduce hydrogen peroxide generated in response to growth factors and tumor necrosis factor-alpha. *J Biol Chem,* 273**,** 6297-302.
KANG, S., CHANG, T., LEE, T., KIM, E., YU, D. & RHEE, S. 2004. Cytosolic peroxiredoxin attenuates the activation of Jnk and p38 but potentiates that of Erk in Hela cells stimulated with tumor necrosis factor-alpha. *J Biol Chem,* 279**,** 2535-43.
KAWANO, Y. & KYPTA, R. 2003. Secreted antagonists of the Wnt signalling pathway. *J Cell Sci,* 116**,** 2627-34.
KELSH, R. & EISEN, J. 2000. The zebrafish colourless gene regulates development of non-ectomesenchymal neural crest derivatives. *Development,* 127**,** 515-25.
KELSH, R. N. 2004. Genetics and evolution of pigment patterns in fish. *Pigment Cell Res,* 17**,** 326-36.
KELSH, R. N. 2006. Sorting out Sox10 functions in neural crest development. *Bioessays,* 28**,** 788-798.
KELSH, R. N., BRAND, M., JIANG, Y. J., HEISENBERG, C. P., LIN, S., HAFFTER, P., ODENTHAL, J., MULLINS, M. C., VAN EEDEN, F. J., FURUTANI-SEIKI, M., GRANATO, M., HAMMERSCHMIDT, M., KANE, D. A., WARGA, R. M., BEUCHLE, D., VOGELSANG, L. & NUSSLEIN-VOLHARD, C. 1996. Zebrafish pigmentation mutations and the processes of neural crest development. *Development,* 123**,** 369-89.
KELSH, R., DUTTON, K., MEDLIN, J. & EISEN, J. 2000. Expression of zebrafish fkd6 in neural crest-derived glia. *Mech Dev,* 93**,** 161-4.
KELSH, R., HARRIS, M., COLANESI, S. & ERICKSON, C. 2009. Stripes and belly-spots -- a review of pigment cell morphogenesis in vertebrates. *Semin Cell Dev Biol,* 20**,** 90-104.
KIL, S., LALLIER, T. & BRONNER-FRASER, M. 1996. Inhibition of cranial neural crest adhesion in vitro and migration in vivo using integrin antisense oligonucleotides. *Dev Biol,* 179**,** 91-101.
KIM, J., LO, L., DORMAND, E. & ANDERSON, D. J. 2003. SOX10 maintains multipotency and inhibits neuronal differentiation of neural crest stem cells. *Neuron,* 38**,** 17-31.
KIMMEL, C. B., BALLARD, W. W., KIMMEL, S. R., ULLMANN, B. & SCHILLING, T. F. 1995. Stages of embryonic development of the zebrafish. *Dev Dyn,* 203**,** 253-310.
KNAUT, H., WERZ, C., GEISLER, R. & NÜSSLEIN-VOLHARD, C. 2003. A zebrafish homologue of the chemokine receptor Cxcr4 is a germ-cell guidance receptor. *Nature,* 421**,** 279-82.
KOJIMA, A., NAKAHAMA, K., OHNO-MATSUI, K., SHIMADA, N., MORI, K., ISEKI, S., SATO, T., MOCHIZUKI, M. & MORITA, I. 2008. Connexin 43 contributes to differentiation of retinal pigment epithelial cells via cyclic AMP signaling. *Biochem Biophys Res Commun,* 366**,** 532-8.
KRULL, C., LANSFORD, R., GALE, N., COLLAZO, A., MARCELLE, C., YANCOPOULOS, G., FRASER, S. & BRONNER-FRASER, M. 1997. Interactions of Eph-related receptors and ligands confer rostrocaudal pattern to trunk neural crest migration. *Curr Biol,* 7**,** 571-80.
KUHLBRODT, K., HERBARTH, B., SOCK, E., HERMANS-BORGMEYER, I. & WEGNER, M. 1998. Sox10, a novel transcriptional modulator in glial cells. *J Neurosci,* 18**,** 237-50.
KUPHAL, S., PALM, H., POSER, I. & BOSSERHOFF, A. 2005. Snail-regulated genes in malignant melanoma. *Melanoma Res,* 15**,** 305-13.
KURIYAMA, S. & MAYOR, R. 2008. Molecular analysis of neural crest migration. *Philos Trans R Soc Lond B Biol Sci,* 363**,** 1349-62.
KWAN, K., FUJIMOTO, E., GRABHER, C., MANGUM, B., HARDY, M., CAMPBELL, D., PARANT, J., YOST, H., KANKI, J. & CHIEN, C. 2007. The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. *Dev Dyn,* 236**,** 3088-99.
LABONNE, C. & BRONNER-FRASER, M. 1998. Neural crest induction in Xenopus: evidence for a two-signal model. *Development,* 125**,** 2403-14.
LACOSTA, A., CANUDAS, J., GONZALEZ, C., MUNIESA, P., SARASA, M. & DOMINGUEZ, L. 2007. Pax7 identifies neural crest, chromatophore lineages and pigment stem cells during zebrafish development. *Int J Dev Biol,* 51**,** 327-31.
LANG, D., CHEN, F., MILEWSKI, R., LI, J., LU, M. & EPSTEIN, J. 2000. Pax3 is required for enteric ganglia formation and functions with Sox10 to modulate expression of c-ret. *J Clin Invest,* 106**,** 963-71.
LANG, D., LU, M., HUANG, L., ENGLEKA, K., ZHANG, M., CHU, E., LIPNER, S., SKOULTCHI, A., MILLAR, S. & EPSTEIN, J. 2005. Pax3 functions at a nodal point in melanocyte stem cell differentiation. *Nature,* 433**,** 884-7.
LASSEN, N., ESTEY, T., TANGUAY, R., PAPPA, A., REIMERS, M. & VASILIOU, V. 2005. Molecular cloning, baculovirus expression, and tissue distribution of the zebrafish aldehyde dehydrogenase 2. *Drug Metab Dispos,* 33**,** 649-56.
LATCHMAN, D. 1997. Transcription factors: an overview. *Int J Biochem Cell Biol,* 29**,** 1305-12.

LEE, H., HUANG, H., LIN, C., CHEN, Y. & TSAI, H. 2006. Foxd3 mediates zebrafish myf5 expression during early somitogenesis. *Dev Biol,* 290**,** 359-72.
LEE, K., NAM, S., CHO, E., SEONG, I., LIMB, J., LEE, S. & KIM, J. 2008. Identification of direct regulatory targets of the transcription factor Sox10 based on function and conservation. *BMC Genomics,* 9**,** 408.
LEFEBVRE, V., DUMITRIU, B., PENZO-MÉNDEZ, A., HAN, Y. & PALLAVI, B. 2007. Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. *Int J Biochem Cell Biol,* 39**,** 2195-214.
LEFEBVRE, V., HUANG, W., HARLEY, V., GOODFELLOW, P. & DE CROMBRUGGHE, B. 1997. SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. *Mol Cell Biol,* 17**,** 2336-46.
LEUNG, Y., MA, P., LINK, B. & DOWLING, J. 2008. Factorial microarray analysis of zebrafish retinal development. *Proc Natl Acad Sci U S A,* 105**,** 12909-14.
LEVY, C., KHALED, M. & FISHER, D. E. 2006. MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol Med*.
LEWIS, J., BONNER, J., MODRELL, M., RAGLAND, J., MOON, R., DORSKY, R. & RAIBLE, D. 2004. Reiterated Wnt signaling during zebrafish neural crest development. *Development,* 131**,** 1299-308.
LI, S., TONG, Y., XIE, X., WANG, Q., ZHOU, H., HAN, Y., ZHANG, Z., GAO, W., LI, S., ZHANG, X. & BI, R. 2007. Octameric structure of the human bifunctional enzyme PAICS in purine biosynthesis. *J Mol Biol,* 366**,** 1603-14.
LI, Y., RANKIN, S., SINNER, D., KENNY, A., KRIEG, P. & ZORN, A. 2008. Sfrp5 coordinates foregut specification and morphogenesis by antagonizing both canonical and noncanonical Wnt11 signaling. *Genes Dev,* 22**,** 3050-63.
LI, Y., ZDANOWICZ, M., YOUNG, L., KUMISKI, D., LEATHERBURY, L. & KIRBY, M. 2003. Cardiac neural crest in zebrafish embryos contributes to myocardial cell lineage and early heart function. *Dev Dyn,* 226**,** 540-50.
LISTER, J., CLOSE, J. & RAIBLE, D. 2001. Duplicate mitf genes in zebrafish: complementary expression and conservation of melanogenic potential. *Dev Biol,* 237**,** 333-44.
LISTER, J., COOPER, C., NGUYEN, K., MODRELL, M., GRANT, K. & RAIBLE, D. 2006. Zebrafish Foxd3 is required for development of a subset of neural crest derivatives. *Dev Biol,* 290**,** 92-104.
LISTER, J., ROBERTSON, C., LEPAGE, T., JOHNSON, S. & RAIBLE, D. 1999. nacre encodes a zebrafish microphthalmia-related protein that regulates neural-crest-derived pigment cell fate. *Development,* 126**,** 3757-67.
LIU, Q., MELNIKOVA, I., HU, M. & GARDNER, P. 1999. Cell type-specific activation of neuronal nicotinic acetylcholine receptor subunit genes by Sox10. *J Neurosci,* 19**,** 9747-55.
LO, C., COHEN, M., HUANG, G., LAZATIN, B., PATEL, N., SULLIVAN, R., PAUKEN, C. & PARK, S. 1997. Cx43 gap junction gene expression and gap junctional communication in mouse neural crest cells. *Dev Genet,* 20**,** 119-32.
LO, L., TIVERON, M. & ANDERSON, D. 1998. MASH1 activates expression of the paired homeodomain transcription factor Phox2a, and couples pan-neuronal and subtype-specific components of autonomic neuronal identity. *Development,* 125**,** 609-20.
LOPES, S., YANG, X., MÜLLER, J., CARNEY, T., MCADOW, A., RAUCH, G., JACOBY, A., HURST, L., DELFINO-MACHÍN, M., HAFFTER, P., GEISLER, R., JOHNSON, S., WARD, A. & KELSH, R. 2008. Leukocyte tyrosine kinase functions in pigment cell development. *PLoS Genet,* 4**,** e1000026.
LUDWIG, A., REHBERG, S. & WEGNER, M. 2004. Melanocyte-specific expression of dopachrome tautomerase is dependent on synergistic gene activation by the Sox10 and Mitf transcription factors. *FEBS Lett,* 556**,** 236-44.
LUO, C., PISARSKA, M. & HSUEH, A. 2005. Identification of a stanniocalcin paralog, stanniocalcin-2, in fish and the paracrine actions of stanniocalcin-2 in the mammalian ovary. *Endocrinology,* 146**,** 469-76.
MA, Q., FODE, C., GUILLEMOT, F. & ANDERSON, D. 1999. Neurogenin1 and neurogenin2 control two distinct waves of neurogenesis in developing dorsal root ganglia. *Genes Dev,* 13**,** 1717-28.
MADERSPACHER, F. & NÜSSLEIN-VOLHARD, C. 2003. Formation of the adult pigment pattern in zebrafish requires leopard and obelix dependent cell interactions. *Development,* 130**,** 3447-57.
MALICKI, J., SCHIER, A., SOLNICA-KREZEL, L., STEMPLE, D., NEUHAUSS, S., STAINIER, D., ABDELILAH, S., RANGINI, Z., ZWARTKRUIS, F. & DRIEVER, W. 1996. Mutations affecting development of the zebrafish ear. *Development,* 123**,** 275-83.

MANANDHAR, G., MIRANDA-VIZUETE, A., PEDRAJAS, J., KRAUSE, W., ZIMMERMAN, S., SUTOVSKY, M. & SUTOVSKY, P. 2009. Peroxiredoxin 2 and Peroxidase Enzymatic Activity of Mammalian Spermatozoa. *Biol Reprod*.

MANSOURI, A., STOYKOVA, A., TORRES, M. & GRUSS, P. 1996. Dysgenesis of cephalic neural crest derivatives in Pax7-/- mutant mice. *Development,* 122**,** 831-8.

MARCHANT, L., LINKER, C., RUIZ, P., GUERRERO, N. & MAYOR, R. 1998. The inductive properties of mesoderm suggest that the neural crest cells are specified by a BMP gradient. *Dev Biol,* 198**,** 319-29.

MATSUKAWA, T., SUGITANI, K., MAWATARI, K., KORIYAMA, Y., LIU, Z., TANAKA, M. & KATO, S. 2004. Role of purpurin as a retinol-binding protein in goldfish retina during the early stage of optic nerve regeneration: its priming action on neurite outgrowth. *J Neurosci,* 24**,** 8346-53.

MATTHEWS, H., BRODERS-BONDON, F., THIERY, J. & MAYOR, R. 2008. Wnt11r is required for cranial neural crest migration. *Dev Dyn,* 237**,** 3404-9.

MERTIN, S., MCDOWALL, S. & HARLEY, V. 1999. The DNA-binding specificity of SOX9 and other SOX proteins. *Nucleic Acids Res,* 27**,** 1359-64.

MEŞE, G., RICHARD, G. & WHITE, T. 2007. Gap junctions: basic structure and function. *J Invest Dermatol,* 127**,** 2516-24.

MEULEMANS, D. & BRONNER-FRASER, M. 2004. Gene-regulatory interactions in neural crest evolution and development. *Dev Cell,* 7**,** 291-9.

MILLENAAR, F., OKYERE, J., MAY, S., VAN ZANTEN, M., VOESENEK, L. & PEETERS, A. 2006. How to decide? Different methods of calculating gene expression from short oligonucleotide array data will give different results. *BMC Bioinformatics,* 7**,** 137.

MINCHIN, J. & HUGHES, S. 2008. Sequential actions of Pax3 and Pax7 drive xanthophore development in zebrafish neural crest. *Dev Biol,* 317**,** 508-22.

MIYASAKA, N., KNAUT, H. & YOSHIHARA, Y. 2007. Cxcl12/Cxcr4 chemokine signaling is required for placode assembly and sensory axon pathfinding in the zebrafish olfactory system. *Development,* 134**,** 2459-68.

MONSORO-BURQ, A., WANG, E. & HARLAND, R. 2005. Msx1 and Pax3 cooperate to mediate FGF8 and WNT signals during Xenopus neural crest induction. *Dev Cell,* 8**,** 167-78.

MONTANER, D., TÁRRAGA, J., HUERTA-CEPAS, J., BURGUET, J., VAQUERIZAS, J., CONDE, L., MINGUEZ, P., VERA, J., MUKHERJEE, S., VALLS, J., PUJANA, M., ALLOZA, E., HERRERO, J., AL-SHAHROUR, F. & DOPAZO, J. 2006. Next station in microarray data analysis: GEPAS. *Nucleic Acids Res,* 34**,** W486-91.

MONTERO-BALAGUER, M., LANG, M., SACHDEV, S., KNAPPMEYER, C., STEWART, R., DE LA GUARDIA, A., HATZOPOULOS, A. & KNAPIK, E. 2006. The mother superior mutation ablates foxd3 activity in neural crest progenitor cells and depletes neural crest derivatives in zebrafish. *Dev Dyn,* 235**,** 3199-212.

MORIN, X., CREMER, H., HIRSCH, M., KAPUR, R., GORIDIS, C. & BRUNET, J. 1997. Defects in sensory and autonomic ganglia and absence of locus coeruleus in mice deficient for the homeobox gene Phox2a. *Neuron,* 18**,** 411-23.

MORRIS, S., NAEVE, C., MATHEW, P., JAMES, P., KIRSTEIN, M., CUI, X. & WITTE, D. 1997. ALK, the chromosome 2 gene locus altered by the t(2;5) in non-Hodgkin's lymphoma, encodes a novel neural receptor tyrosine kinase that is highly related to leukocyte tyrosine kinase (LTK). *Oncogene,* 14**,** 2175-88.

MORRISON, S., PEREZ, S., QIAO, Z., VERDI, J., HICKS, C., WEINMASTER, G. & ANDERSON, D. 2000. Transient Notch activation initiates an irreversible switch from neurogenesis to gliogenesis by neural crest stem cells. *Cell,* 101**,** 499-510.

MURAYAMA, E., HERBOMEL, P., KAWAKAMI, A., TAKEDA, H. & NAGASAWA, H. 2005. Otolith matrix proteins OMP-1 and Otolin-1 are necessary for normal otolith growth and their correct anchoring onto the sensory maculae. *Mech Dev,* 122**,** 791-803.

MURISIER, F. & BEERMANN, F. 2006. Genetics of pigment cells: lessons from the tyrosinase gene family. *Histol Histopathol,* 21**,** 567-78.

MURISIER, F., GUICHARD, S. & BEERMANN, F. 2006. A conserved transcriptional enhancer that specifies Tyrp1 expression to melanocytes. *Dev Biol,* 298**,** 644-55.

MURISIER, F., GUICHARD, S. & BEERMANN, F. 2007. The tyrosinase enhancer is activated by Sox10 and Mitf in mouse melanocytes. *Pigment Cell Res,* 20**,** 173-84.

NATARAJAN, D., MARCOS-GUTIERREZ, C., PACHNIS, V. & DE GRAAFF, E. 2002. Requirement of signalling by receptor tyrosine kinase RET for the directed migration of enteric nervous system progenitor cells during mammalian embryogenesis. *Development,* 129**,** 5151-60.

NGUYEN, V., SCHMID, B., TROUT, J., CONNORS, S., EKKER, M. & MULLINS, M. 1998. Ventral and lateral regions of the zebrafish gastrula, including the neural crest progenitors, are established by a bmp2b/swirl pathway of genes. *Dev Biol,* 199**,** 93-110.

NISHI, T. & FORGAC, M. 2002. The vacuolar (H+)-ATPases--nature's most versatile proton pumps. *Nat Rev Mol Cell Biol,* 3**,** 94-103.

NOBUKUNI, Y., WATANABE, A., TAKEDA, K., SKARKA, H. & TACHIBANA, M. 1996. Analyses of loss-of-function mutations of the MITF gene suggest that haploinsufficiency is a cause of Waardenburg syndrome type 2A. *Am J Hum Genet,* 59**,** 76-83.

ODENTHAL, J. & NÜSSLEIN-VOLHARD, C. 1998. fork head domain genes in zebrafish. *Dev Genes Evol,* 208**,** 245-58.

ODENTHAL, J., ROSSNAGEL, K., HAFFTER, P., KELSH, R., VOGELSANG, E., BRAND, M., VAN EEDEN, F., FURUTANI-SEIKI, M., GRANATO, M., HAMMERSCHMIDT, M., HEISENBERG, C., JIANG, Y., KANE, D., MULLINS, M. & NÜSSLEIN-VOLHARD, C. 1996. Mutations affecting xanthophore pigmentation in the zebrafish, Danio rerio. *Development,* 123**,** 391-8.

O'DONNELL, M., HONG, C., HUANG, X., DELNICKI, R. & SAINT-JEANNET, J. 2006. Functional analysis of Sox8 during neural crest development in Xenopus. *Development,* 133**,** 3817-26.

OPDECAMP, K., NAKAYAMA, A., NGUYEN, M., HODGKINSON, C., PAVAN, W. & ARNHEITER, H. 1997. Melanocyte development in vivo and in neural crest cell cultures: crucial dependence on the Mitf basic-helix-loop-helix-zipper transcription factor. *Development,* 124**,** 2377-86.

OSAWA, M., EGAWA, G., MAK, S., MORIYAMA, M., FRETER, R., YONETANI, S., BEERMANN, F. & NISHIKAWA, S. 2005. Molecular characterization of melanocyte stem cells in their niche. *Development,* 132**,** 5589-99.

PARATORE, C., EICHENBERGER, C., SUTER, U. & SOMMER, L. 2002. Sox10 haploinsufficiency affects maintenance of progenitor cells in a mouse model of Hirschsprung disease. *Hum Mol Genet,* 11**,** 3075-85.

PARATORE, C., GOERICH, D., SUTER, U., WEGNER, M. & SOMMER, L. 2001. Survival and glial fate acquisition of neural crest cells are regulated by an interplay between the transcription factor Sox10 and extrinsic combinatorial signaling. *Development,* 128**,** 3949-61.

PARICHY, D., RAWLS, J., PRATT, S., WHITFIELD, T. & JOHNSON, S. 1999. Zebrafish sparse corresponds to an orthologue of c-kit and is required for the morphogenesis of a subpopulation of melanocytes, but is not essential for hematopoiesis or primordial germ cell development. *Development,* 126**,** 3425-36.

PATTYN, A., MORIN, X., CREMER, H., GORIDIS, C. & BRUNET, J. 1999. The homeobox gene Phox2b is essential for the development of autonomic neural crest derivatives. *Nature,* 399**,** 366-70.

PEIRANO, R., GOERICH, D., RIETHMACHER, D. & WEGNER, M. 2000. Protein zero gene expression is regulated by the glial transcription factor Sox10. *Mol Cell Biol,* 20**,** 3198-209.

PELLETIER, I., BALLY-CUIF, L. & ZIEGLER, I. 2001. Cloning and developmental expression of zebrafish GTP cyclohydrolase I. *Mech Dev,* 109**,** 99-103.

PEREZ, S., REBELO, S. & ANDERSON, D. 1999. Early specification of sensory neuron fate revealed by expression and function of neurogenins in the chick embryo. *Development,* 126**,** 1715-28.

PFEIFER, D., POULAT, F., HOLINSKI-FEDER, E., KOOY, F. & SCHERER, G. 2000. The SOX8 gene is located within 700 kb of the tip of chromosome 16p and is deleted in a patient with ATR-16 syndrome. *Genomics,* 63**,** 108-16.

PICKART, M. A., SIVASUBBU, S., NIELSEN, A. L., SHRIRAM, S., KING, R. A. & EKKER, S. C. 2004. Functional genomics tools for the analysis of zebrafish pigment. *Pigment Cell Res,* 17**,** 461-70.

PIETREMENT, C., SUN-WADA, G., SILVA, N., MCKEE, M., MARSHANSKY, V., BROWN, D., FUTAI, M. & BRETON, S. 2006. Distinct expression patterns of different subunit isoforms of the V-ATPase in the rat epididymis. *Biol Reprod,* 74**,** 185-94.

PINGAULT, V., BONDURAND, N., KUHLBRODT, K., GOERICH, D., PRÉHU, M., PULITI, A., HERBARTH, B., HERMANS-BORGMEYER, I., LEGIUS, E., MATTHIJS, G., AMIEL, J., LYONNET, S., CECCHERINI, I., ROMEO, G., SMITH, J., READ, A., WEGNER, M. & GOOSSENS, M. 1998. SOX10 mutations in patients with Waardenburg-Hirschsprung disease. *Nat Genet,* 18**,** 171-3.

POTTERF, S. B., FURUMURA, M., DUNN, K. J., ARNHEITER, H. & PAVAN, W. J. 2000. Transcription factor hierarchy in Waardenburg syndrome: regulation of MITF expression by SOX10 and PAX3. *Hum Genet,* 107**,** 1-6.

POTTERF, S. B., MOLLAAGHABABA, R., HOU, L., SOUTHARD-SMITH, E. M., HORNYAK, T. J., ARNHEITER, H. & PAVAN, W. J. 2001. Analysis of SOX10 function in neural crest-derived melanocyte development: SOX10-dependent transcriptional control of dopachrome tautomerase. *Dev Biol,* 237**,** 245-57.

QIAN, F., ZHEN, F., ONG, C., JIN, S. W., MENG SOO, H., STAINIER, D. Y., LIN, S., PENG, J. & WEN, Z. 2005. Microarray analysis of zebrafish cloche mutant using amplified cDNA and identification of potential downstream target genes. *Dev Dyn,* 233**,** 1163-72.

RABIONET, R., LÓPEZ-BIGAS, N., ARBONÈS, M. & ESTIVILL, X. 2002. Connexin mutations in hearing loss, dermatological and neurological disorders. *Trends Mol Med,* 8**,** 205-12.

RAIBLE, D. & EISEN, J. 1994. Restriction of neural crest cell fate in the trunk of the embryonic zebrafish. *Development,* 120**,** 495-503.

RAIBLE, D. W., WOOD, A., HODSDON, W., HENION, P. D., WESTON, J. A. & EISEN, J. S. 1992. Segregation and early dispersal of neural crest cells in the embryonic zebrafish. *Dev Dyn,* 195**,** 29-42.

RAYL, E., MOROSON, B. & BEARDSLEY, G. 1996. The human purH gene product, 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase. Cloning, sequencing, expression, purification, kinetic analysis, and domain mapping. *J Biol Chem,* 271**,** 2225-33.

REEVES, N. & POSAKONY, J. W. 2005. Genetic programs activated by proneural proteins in the developing Drosophila PNS. *Dev Cell,* 8**,** 413-25.

REMÉNYI, A., LINS, K., NISSEN, L., REINBOLD, R., SCHÖLER, H. & WILMANNS, M. 2003. Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev,* 17**,** 2048-59.

ROHRLICH, S. & RUBIN, R. 1975. Biochemical characterization of crystals from the dermal iridophores of a chameleon Anolis carolinensis. *J Cell Biol,* 66**,** 635-45.

ROSENTRETER, A., HOFMANN, A., XAVIER, C., STUMPF, M., NOEGEL, A. & CLEMEN, C. 2007. Coronin 3 involvement in F-actin-dependent processes at the cell cortex. *Exp Cell Res,* 313**,** 878-95.

SAEKI, H. & OIKAWA, A. 1978. Effects of pH and type of sugar in the medium on tyrosinase activity in cultured melanoma cells. *J Cell Physiol,* 94**,** 139-45.

SANTIAGO, A. & ERICKSON, C. 2002. Ephrin-B ligands play a dual role in the control of neural crest cell migration. *Development,* 129**,** 3621-32.

SAUKA-SPENGLER, T. & BRONNER-FRASER, M. 2008. A gene regulatory network orchestrates neural crest formation. *Nat Rev Mol Cell Biol,* 9**,** 557-68.

SCAFFIDI, P. & BIANCHI, M. 2001. Spatially precise DNA bending is an essential activity of the sox2 transcription factor. *J Biol Chem,* 276**,** 47296-302.

SCHALLREUTER, K., KOTHARI, S., CHAVAN, B. & SPENCER, J. 2008. Regulation of melanogenesis--controversies and new concepts. *Exp Dermatol,* 17**,** 395-404.

SCHEPERS, G., BULLEJOS, M., HOSKING, B. & KOOPMAN, P. 2000. Cloning and characterisation of the Sry-related transcription factor gene Sox8. *Nucleic Acids Res,* 28**,** 1473-80.

SCHEPERS, G., TEASDALE, R. & KOOPMAN, P. 2002. Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. *Dev Cell,* 3**,** 167-70.

SCHLIERF, B., WERNER, T., GLASER, G. & WEGNER, M. 2006. Expression of connexin47 in oligodendrocytes is regulated by the Sox10 transcription factor. *J Mol Biol,* 361**,** 11-21.

SCHMIDT, K., SCHINKE, T., HABERLAND, M., PRIEMEL, M., SCHILLING, A., MUELDNER, C., RUEGER, J., SOCK, E., WEGNER, M. & AMLING, M. 2005. The high mobility group transcription factor Sox8 is a negative regulator of osteoblast differentiation. *J Cell Biol,* 168**,** 899-910.

SCHUBERT, D. & LACORBIERE, M. 1985. Isolation of an adhesion-mediating protein from chick neural retina adherons. *J Cell Biol,* 101**,** 1071-7.

SCHUBERT, D., LACORBIERE, M. & ESCH, F. 1986. A chick neural retina adhesion and survival molecule is a retinol-binding protein. *J Cell Biol,* 102**,** 2295-301.

SERBEDZIJA, G., BRONNER-FRASER, M. & FRASER, S. 1994. Developmental potential of trunk neural crest cells in the mouse. *Development,* 120**,** 1709-18.

SHAH, N. & ANDERSON, D. 1997. Integration of multiple instructive cues by neural crest stem cells reveals cell-intrinsic biases in relative growth factor responsiveness. *Proc Natl Acad Sci U S A,* 94**,** 11369-74.

SHAPIRO, H. M. 2003. *Practical Flow Cytometry,* Hoboken, NJ., John Wiley & Sons Inc.

SHEN, C. & NATHAN, C. 2002. Nonredundant antioxidant defense by multiple two-cysteine peroxiredoxins in human prostate cancer cells. *Mol Med,* 8**,** 95-102.

SHOOK, D. & KELLER, R. 2003. Mechanisms, mechanics and function of epithelial-mesenchymal transitions in early development. *Mech Dev,* 120**,** 1351-83.

SIEBER-BLUM, M., GRIM, M., HU, Y. & SZEDER, V. 2004. Pluripotent neural crest stem cells in the adult hair follicle. *Dev Dyn,* 231**,** 258-69.

SOCK, E., SCHMIDT, K., HERMANNS-BORGMEYER, I., BÖSL, M. & WEGNER, M. 2001. Idiopathic weight reduction in mice deficient in the high-mobility-group transcription factor Sox8. *Mol Cell Biol,* 21**,** 6951-9.

SOMMER, L., SHAH, N., RAO, M. & ANDERSON, D. 1995. The cellular function of MASH1 in autonomic neurogenesis. *Neuron,* 15**,** 1245-58.

SONG, W., ZOU, Z., XU, F., GU, X., XU, X. & ZHAO, Q. 2006. Molecular cloning and expression of a second zebrafish aldehyde dehydrogenase 2 gene (aldh2b). *DNA Seq,* 17**,** 262-9.

SOUTHARD-SMITH, E., KOS, L. & PAVAN, W. 1998. Sox10 mutation disrupts neural crest development in Dom Hirschsprung mouse model. *Nat Genet,* 18**,** 60-4.

SPRAGUE, J., BAYRAKTAROGLU, L., CLEMENTS, D., CONLIN, T., FASHENA, D., FRAZER, K., HAENDEL, M., HOWE, D., MANI, P., RAMACHANDRAN, S., SCHAPER, K., SEGERDELL, E., SONG, P., SPRUNGER, B., TAYLOR, S., VAN SLYKE, C. & WESTERFIELD, M. 2006. The Zebrafish Information Network: the zebrafish model organism database. *Nucleic Acids Res,* 34**,** D581-5.

STEMPLE, D. & ANDERSON, D. 1992. Isolation of a stem cell for neurons and glia from the mammalian neural crest. *Cell,* 71**,** 973-85.

STEVENTON, B., CARMONA-FONTAINE, C. & MAYOR, R. 2005. Genetic network during neural crest induction: from cell specification to cell survival. *Semin Cell Dev Biol,* 16**,** 647-54.

STEWART, R., ARDUINI, B., BERGHMANS, S., GEORGE, R., KANKI, J., HENION, P. & LOOK, A. 2006. Zebrafish foxd3 is selectively required for neural crest specification, migration and survival. *Dev Biol,* 292**,** 174-88.

STOLT, C., LOMMES, P., FRIEDRICH, R. & WEGNER, M. 2004. Transcription factors Sox8 and Sox10 perform non-equivalent roles during oligodendrocyte development despite functional redundancy. *Development,* 131**,** 2349-58.

STOREY, J. & TIBSHIRANI, R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A,* 100**,** 9440-5.

STOUGHTON, R. B. 2005. Applications of DNA microarrays in biology. *Annu Rev Biochem,* 74**,** 53-82.

STOUT, C., GOODENOUGH, D. & PAUL, D. 2004. Connexins: functions without junctions. *Curr Opin Cell Biol,* 16**,** 507-12.

SUMANAS, S., JORNIAK, T. & LIN, S. 2005. Identification of novel vascular endothelial-specific genes by the microarray analysis of the zebrafish cloche mutants. *Blood,* 106**,** 534-41.

SVETIC, V., HOLLWAY, G. E., ELWORTHY, S., CHIPPERFIELD, T. R., DAVISON, C., ADAMS, R. J., EISEN, J. S., INGHAM, P. W., CURRIE, P. D. & KELSH, R. N. 2007. Sdf1a patterns zebrafish melanophores and links the somite and melanophore pattern defects in choker mutants. *Development,* 134**,** 1011-1022.

SYRRIS, P., CARTER, N. & PATTON, M. 1999. Novel nonsense mutation of the endothelin-B receptor gene in a family with Waardenburg-Hirschsprung disease. *American Journal of Medical Genetics,* 87**,** 69-71.

TABATA, H., KAWAMURA, N., SUN-WADA, G. & WADA, Y. 2008. Vacuolar-type H(+)-ATPase with the a3 isoform is the proton pump on premature melanosomes. *Cell Tissue Res,* 332**,** 447-60.

TACHIBANA, M. 1999. Sound needs sound melanocytes to be heard. *Pigment Cell Res,* 12**,** 344-54.

TACHIBANA, M., KOBAYASHI, Y. & MATSUSHIMA, Y. 2003. Mouse models for four types of Waardenburg syndrome. *Pigment Cell Res,* 16**,** 448-54.

TANAKA, M., MURAYAMA, D., NAGASHIMA, M., HIGASHI, T., MAWATARI, K., MATSUKAWA, T. & KATO, S. 2007. Purpurin expression in the zebrafish retina during early development and after optic nerve lesion in adults. *Brain Res,* 1153**,** 34-42.

TANAKA, T., JARADAT, S., LIM, M., KARGUL, G., WANG, X., GRAHOVAC, M., PANTANO, S., SANO, Y., PIAO, Y., NAGARAJA, R., DOI, H., WOOD, W. R., BECKER, K. & KO, M. 2000. Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. *Proc Natl Acad Sci U S A,* 97**,** 9127-32.

TANEYHILL, L. 2008. To adhere or not to adhere: the role of Cadherins in neural crest development. *Cell Adh Migr,* 2**,** 223-30.

TÁRRAGA, J., MEDINA, I., CARBONELL, J., HUERTA-CEPAS, J., MINGUEZ, P., ALLOZA, E., AL-SHAHROUR, F., VEGAS-AZCÁRATE, S., GOETZ, S., ESCOBAR, P., GARCIA-GARCIA, F., CONESA, A., MONTANER, D. & DOPAZO, J. 2008. GEPAS, a web-based tool for microarray data analysis and interpretation. *Nucleic Acids Res,* 36**,** W308-14.

TASSABEHJI, M., NEWTON, V. & READ, A. 1994. Waardenburg syndrome type 2 caused by mutations in the human microphthalmia (MITF) gene. *Nat Genet,* 8**,** 251-5.

TASSABEHJI, M., READ, A., NEWTON, V., HARRIS, R., BALLING, R., GRUSS, P. & STRACHAN, T. 1992. Waardenburg's syndrome patients have mutations in the human homologue of the Pax-3 paired box gene. *Nature,* 355**,** 635-6.

TESTAZ, S., DELANNET, M. & DUBAND, J. 1999. Adhesion and migration of avian neural crest cells on fibronectin require the cooperating activities of multiple integrins of the (beta)1 and (beta)3 families. *J Cell Sci,* 112 (Pt 24)**,** 4715-28.

THAL, D., XAVIER, C., ROSENTRETER, A., LINDER, S., FRIEDRICHS, B., WAHA, A., PIETSCH, T., STUMPF, M., NOEGEL, A. & CLEMEN, C. 2008. Expression of coronin-3 (coronin-1C) in diffuse gliomas is related to malignancy. *J Pathol,* 214**,** 415-24.

THIERY, J. & SLEEMAN, J. 2006. Complex networks orchestrate epithelial-mesenchymal transitions. *Nat Rev Mol Cell Biol,* 7**,** 131-42.

THISSE, B., HEYER, V., LUX, A., ALUNNI, V., DEGRAVE, A., SEILIEZ, I., KIRCHNER, J., PARKHILL, J. & THISSE, C. 2004. Spatial and temporal expression of the zebrafish genome by large-scale in situ hybridization screening. *Methods Cell Biol,* 77**,** 505-19.

THOMAS, A. & ERICKSON, C. 2008. The making of a melanocyte: the specification of melanoblasts from the neural crest. *Pigment Cell Melanoma Res,* 21**,** 598-610.

TRIBULO, C., AYBAR, M., NGUYEN, V., MULLINS, M. & MAYOR, R. 2003. Regulation of Msx genes by a Bmp gradient is essential for neural crest specification. *Development,* 130**,** 6441-52.

TRINEI, M., LANFRANCONE, L., CAMPO, E., PULFORD, K., MASON, D., PELICCI, P. & FALINI, B. 2000. A new variant anaplastic lymphoma kinase (ALK)-fusion protein (ATIC-ALK) in a case of ALK-positive anaplastic large cell lymphoma. *Cancer Res,* 60**,** 793-8.

TSUDA, M., TAKAHASHI, S., TAKAHASHI, Y. & ASAHARA, H. 2003. Transcriptional co-activators CREB-binding protein and p300 regulate chondrocyte-specific gene expression via association with Sox9. *J Biol Chem,* 278**,** 27224-9.

UCHIKAWA, M., KAMACHI, Y. & KONDOH, H. 1999. Two distinct subgroups of Group B Sox genes for transcriptional activators and repressors: their expression during embryonic organogenesis of the chicken. *Mech Dev,* 84**,** 103-20.

UETRECHT, A. & BEAR, J. 2006. Coronins: the return of the crown. *Trends Cell Biol,* 16**,** 421-6.

VALLS, J., GRAU, M., SOLÉ, X., HERNÁNDEZ, P., MONTANER, D., DOPAZO, J., PEINADO, M., CAPELLÁ, G., MORENO, V. & PUJANA, M. 2008. CLEAR-test: combining inference for differential expression and variability in microarray data analysis. *J Biomed Inform,* 41**,** 33-45.

VAN EEDEN, F., GRANATO, M., SCHACH, U., BRAND, M., FURUTANI-SEIKI, M., HAFFTER, P., HAMMERSCHMIDT, M., HEISENBERG, C., JIANG, Y., KANE, D., KELSH, R., MULLINS, M., ODENTHAL, J., WARGA, R., ALLENDE, M., WEINBERG, E. & NÜSSLEIN-VOLHARD, C. 1996. Mutations affecting somite formation and patterning in the zebrafish, Danio rerio. *Development,* 123**,** 153-64.

VERASTEGUI, C., BILLE, K., ORTONNE, J. & BALLOTTI, R. 2000. Regulation of the microphthalmia-associated transcription factor gene by the Waardenburg syndrome type 4 gene, SOX10. *J Biol Chem,* 275**,** 30757-60.

VILLEFRANC, J., AMIGO, J. & LAWSON, N. 2007. Gateway compatible vectors for analysis of gene function in the zebrafish. *Dev Dyn,* 236**,** 3077-87.

WAGNER, T., WIRTH, J., MEYER, J., ZABEL, B., HELD, M., ZIMMER, J., PASANTES, J., BRICARELLI, F., KEUTEL, J. & HUSTERT, E. 1994. Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. *Cell,* 79**,** 1111-20.

WANG, Q., KUMAR, S., MITSIOS, N., SLEVIN, M. & KUMAR, P. 2007. Investigation of downstream target genes of PAX3c, PAX3e and PAX3g isoforms in melanocytes by microarray analysis. *Int J Cancer,* 120**,** 1223-31.

WARDLE, F., ODOM, D., BELL, G., YUAN, B., DANFORD, T., WIELLETTE, E., HERBOLSHEIMER, E., SIVE, H., YOUNG, R. & SMITH, J. 2006. Zebrafish promoter microarrays identify actively transcribed embryonic genes. *Genome Biol,* 7**,** R71.

WATANABE, M., IWASHITA, M., ISHII, M., KURACHI, Y., KAWAKAMI, A., KONDO, S. & OKADA, N. 2006. Spot pattern of leopard Danio is caused by mutation in the zebrafish connexin41.8 gene. *EMBO Rep,* 7**,** 893-7.

WEI, Q., MISKIMINS, W. & MISKIMINS, R. 2004. Sox10 acts as a tissue-specific transcription factor enhancing activation of the myelin basic protein gene promoter by p27Kip1 and Sp1. *J Neurosci Res,* 78**,** 796-802.

WHITFIELD, T. T., GRANATO, M., VAN EEDEN, F. J., SCHACH, U., BRAND, M., FURUTANI-SEIKI, M., HAFFTER, P., HAMMERSCHMIDT, M., HEISENBERG, C. P., JIANG, Y. J., KANE, D. A., KELSH, R. N., MULLINS, M. C., ODENTHAL, J. & NUSSLEIN-VOLHARD, C. 1996. Mutations affecting development of the zebrafish inner ear and lateral line. *Development,* 123**,** 241-54.

WILCOX, E. R., BURTON, Q. L., NAZ, S., RIAZUDDIN, S., SMITH, T. N., PLOPLIS, B., BELYANTSEVA, I., BEN-YOSEF, T., LIBURD, N. A., MORELL, R. J., KACHAR, B., WU,

D. K., GRIFFITH, A. J. & FRIEDMAN, T. B. 2001. Mutations in the gene encoding tight junction claudin-14 cause autosomal recessive deafness DFNB29. *Cell,* 104**,** 165-72.
WILSON, M. & KOOPMAN, P. 2002. Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. *Curr Opin Genet Dev,* 12**,** 441-6.
WONG, C., PARATORE, C., DOURS-ZIMMERMANN, M., ROCHAT, A., PIETRI, T., SUTER, U., ZIMMERMANN, D., DUFOUR, S., THIERY, J., MEIJER, D., BEERMANN, F., BARRANDON, Y. & SOMMER, L. 2006. Neural crest-derived cells with stem cell features can be traced back to multiple lineages in the adult skin. *J Cell Biol,* 175**,** 1005-15.
WRIGHT, E., HARGRAVE, M., CHRISTIANSEN, J., COOPER, L., KUN, J., EVANS, T., GANGADHARAN, U., GREENFIELD, A. & KOOPMAN, P. 1995. The Sry-related gene Sox9 is expressed during chondrogenesis in mouse embryos. *Nat Genet,* 9**,** 15-20.
XU, X., FRANCIS, R., WEI, C., LINASK, K. & LO, C. 2006. Connexin 43-mediated modulation of polarized cell movement and the directional migration of cardiac neural crest cells. *Development,* 133**,** 3629-39.
XU, X., LI, W., HUANG, G., MEYER, R., CHEN, T., LUO, Y., THOMAS, M., RADICE, G. & LO, C. 2001. Modulation of mouse neural crest cell motility by N-cadherin and connexin 43 gap junctions. *J Cell Biol,* 154**,** 217-30.
YOUNG, H., HEARN, C., FARLIE, P., CANTY, A., THOMAS, P. & NEWGREEN, D. 2001. GDNF is a chemoattractant for enteric neural cells. *Dev Biol,* 229**,** 503-16.
YUAN, H., CORBI, N., BASILICO, C. & DAILEY, L. 1995. Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev,* 9**,** 2635-45.
ZHOU, W., YE, X., SUN, Z., JI, X., CHEN, H. & XIE, D. 2009. Overexpression of degenerative spermatocyte homolog 1 up-regulates the expression of cyclin D1 and enhances metastatic efficiency in esophageal carcinoma Eca109 cells. *Mol Carcinog*.
ZHOU, Y. & SNEAD, M. 2008. Derivation of cranial neural crest-like cells from human embryonic stem cells. *Biochem Biophys Res Commun,* 376**,** 542-7.
ZHU, H., TRAVER, D., DAVIDSON, A. J., DIBIASE, A., THISSE, C., THISSE, B., NIMER, S. & ZON, L. I. 2005. Regulation of the lmo2 promoter during hematopoietic and vascular development in zebrafish. *Dev Biol,* 281**,** 256-69.
ZHU, L., LEE, H., JORDAN, C., CANTRELL, V., SOUTHARD-SMITH, E. & SHIN, M. 2004. Spatiotemporal regulation of endothelin receptor-B by SOX10 in neural crest-derived enteric neuron precursors. *Nat Genet,* 36**,** 732-7.
ZIEGLER, I. 2003. The pteridine pathway in zebrafish: regulation and specification during the determination of neural crest cell-fate. *Pigment Cell Res,* 16**,** 172-82.
ZIEGLER, I., MCDONALD, T., HESSLINGER, C., PELLETIER, I., BOYLE, P. & MCDONALDO, T. 2000. Development of the pteridine pathway in the zebrafish, Danio rerio. *J Biol Chem,* 275**,** 18926-32.
ZLOTOGORA, J., LERER, I., BAR-DAVID, S., ERGAZ, Z. & ABELIOVICH, D. 1995. Homozygosity for Waardenburg syndrome. *Am J Hum Genet,* 56**,** 1173-8.
ZOU, J., BEERMANN, F., WANG, J., KAWAKAMI, K. & WEI, X. 2006. The Fugu tyrp1 promoter directs specific GFP expression in zebrafish: tools to study the RPE and the neural crest-derived melanophores. *Pigment Cell Res,* 19**,** 615-27.

Appendix

Affymetrix GeneChip® Definitions

Definitions were taken or adapted from the Affymetrix GeneChip® Expression Analysis Data Analysis Fundamentals handbook on the Affymetrix website (www.Affymetrix.com).

Array (or Chip): A collection of probes on glass encased in a plastic cartridge.

Background: A measurement of signal intensity caused by auto-fluorescence of the array surface and non-specifically bound target or stain molecules.

Change: A qualitative call indicating an increase (I), marginal increase (MI), no change (NC), marginal decrease (MD) or decrease (D) in transcript level between a control array and an experimental array.

Detection call: A qualitative measurement indicating if as transcript has been detected (Present), not detected (Absent) or marginally detected (Marginal).

Hybridization Controls: Controls added to the samples before hybridization to the array, eg: *bioB*.

Noise (Raw Q): The result of small variations in digitized signals in the scanner as it samples the probe array surface and is measured by examining the pixel to pixel variations in signal intensities.

Probe: A 25-mer oligonucleotide on the surface of an array. Hybridization of a target to the probe is used to generate intensity data for measuring levels of gene expression.

Probe Cell: A square shaped area of the array containing probes of a unique sequence.

Probe pair: MM and PM

Probe Set: A collection of probe cells that interrogates the same sequence.

Signal: A quantitative measurement of the relative abundance of a transcript.

Target: The labeled (biotinylated) and fragmented sample applied to an array for hybridization to a matching probe.

Sectioning Protocol

Embryos fixed in 4 % PFA were dehydrated in 70 % ethanol for 2 hours, 96 % ethanol for 2 hours and 100 % ethanol for 1 hour at room temperature. Pre-infiltration was performed using a 1:1 ratio of 100 % ethanol to infiltration solution A for 2 hours at room temperature. For infiltration, embryos were incubated in infiltration solution A for up to 24 hours at room temperature. Embedding and polymerization were performed by positioning the embryos in embedding solution B in a Histoform mould recess and then incubating for 2 hours at room temperature. To mount the embryo, Histobloc holders were placed into the Histoform mould recess, Technovit 3040 liquid was then poured into the Histobloc holder and allowed to polymerize for 10 minutes.Tissue blocks were then sectioned using a microtome. Cut sections were placed on a glass slide to dry and a cover slip glued on top.

Infiltration Solution A:

Technovit 7100 (100 ml)

Hardener 1 (1 g)

Mix well for 1 minute and then store for up to 4 weeks at 4 °C

Embedding Solution B:

Infiltration solution A (15 parts)

Hardener 2 (1 part)

Mix well for 1 minute

Folder Name	File Name	Description
Thesis Files	ThesisWord.docx	Word 2007 document
	ThesisWord.doc	Word 2003 document
	Thesis.pdf	
	Figures (folder)	.jpg images used in thesis
Excel Spreadsheets	7.2Ttest.xlsx	Excel 2007 7.2sox10:GFP analysis
	7.2Ttest.xls	Excel 2003 7.2sox10:GFP analysis
	4.9Ttest.xlsx	Excel 2007 4.9sox10:GFP analysis
	4.9Ttest.xls	Excel 2003 4.9sox10:GFP analysis
	7.2and4.9analysis.xlsx	Excel 2007 combined data set analysis
	7.2and4.9analysis.xls	Excel 2003 combined data set analysis
	MIAMI files (folder)	Minimum information about a microarray experiment files
Original Array Data	Original Array Data (folder)	Raw microarray data from GCOS, .cel, .dat and .chp files
	GEPAS Files (folder)	GEPAS notepad files generated from expresso and t-rex.

Table 65: Table of electronic files provided on the accompanying CD.